





**EFFICIENT TECHNIQUES FOR THE  
SINGLE-FRAME SUPER-RESOLUTION RECONSTRUCTION  
OF INTENSITY IMAGES**

**Ph.D. Thesis by  
Aydın AKYOL**

**Department : Computer Engineering**

**Programme : Computer Engineering**

**FEBRUARY 2012**



**EFFICIENT TECHNIQUES FOR THE  
SINGLE-FRAME SUPER-RESOLUTION RECONSTRUCTION  
OF INTENSITY IMAGES**

**Ph.D. Thesis by  
Aydın AKYOL  
(504032508)**

**Date of submission : 26 September 2011**

**Date of defence examination : 14 February 2012**

**Supervisor(Chairman) : Prof.Dr. Muhittin GÖKMEN**  
**Members of the Examining Committee : Prof.Dr. Bülent SANKUR (B.U.)**  
**Assoc. Prof.Dr. Zehra ÇATALTEPE (I.T.U.)**  
**Asst. Prof.Dr. Hakan ERDOĞAN (Sabanci U.)**  
**Asst. Prof.Dr. Mustafa KAMAŞAK (I.T.U.)**

**FEBRUARY 2012**



**TEK İMGEDEN  
SÜPER-ÇÖZÜNÜRLÜKLÜ GERİ-ÇATMA AMACIYLA  
GELİŞTİRİLMİŞ ETKİN YÖNTEMLER**

**DOKTORA TEZİ  
Aydın AKYOL  
(504032508)**

**Tezin Enstitüye Verildiği Tarih : 26 Eylül 2011**

**Tezin Savunulduğu Tarih : 14 Şubat 2012**

**Tez Danışmanı : Prof.Dr. Muhittin GÖKMEN  
Diğer Jüri Üyeleri : Prof.Dr. Bülent SANKUR (B.Ü.)  
Assoc. Prof.Dr. Zehra ÇATALTEPE (İ.T.Ü.)  
Asst. Prof.Dr. Hakan ERDOĞAN (Sabancı Ü.)  
Asst. Prof.Dr. Mustafa KAMAŞAK (İ.T.Ü.)**

**ŞUBAT 2012**





“Our true mentor in life is science.”

*Mustafa Kemal ATATÜRK*



## FOREWORD

A few lines of acknowledgment would not be enough to fully express my appreciation for those guided and supported me through this long research period. I have been fortunate to be surrounded by wonderful teachers, family and friends making my life easier, enjoyable and meaningful.

I owe my deepest gratitude to my thesis supervisor Prof. Muhittin Gökmen for his boundless support, guidance, patience and for allowing me to work in my own way. He has taught me not only many concepts in image processing and computer vision, but also the methodology to carry out the research, to develop analytic thinking and to cope with the difficulties by focusing on right targets. It is always a great honor for me to be student of him.

I have been very fortunate to have Prof. Bülent Sankur in my thesis follow-up committee. He is an inspirational person with his high ethical values and with his rigor in research as well as with his vast knowledge in various fields. I am deeply indebted to him for helping me advance my research.

I would like to thank Assoc. Prof. Zehra Çataltepe for being member of the thesis follow-up committee and for her valuable feedbacks and affectionate encouragement.

I appreciate the time that Asst. Prof. Hakan Erdoğan and Asst. Prof. Mustafa Kamaşak, the other members of my defense jury, spent reading this thesis carefully and making valuable suggestions for its improvements.

I would also like to express my sincere gratitude to Prof. Marshall Tappen for hosting me at the Computer Vision Laboratory, University of Central Florida, USA, from August 2007 to March 2008. Any moment that I spent with this brilliant scientist was quite valuable and helped me a lot not only get out of ruts in my research but also explore new ideas.

I specifically thank Prof. Nadia Erdoğan for her ingenuous encouragement helped me keep my motivation always up.

Many thanks to all the members of my department at İTÜ for providing me a productive and cheerful environment during my research.

I would like to thank my friends that I have met during my Ph.D. study for making this long research period productive, easier and highly enjoyable. Many thanks to Yusuf Yaslan, Serap Kırbız, Yusuf Aytar, Bülent Taştan, Ö. Bilal Orhan, Murat Orhun, Pınar S. Bölük, Arman Savran, Cem Demirkır, Sanem-Çağatay Talay, Koray Kayabol, Mighty Itauma and Kegan G. G. Samuel.

I also wish to thank other really wonderful friends for their sincere friendship, support and encouragement in this long journey. Many thanks to Gülin Çakmak, Erdem-Erkan Özkan, Demet-Erkan Özkan, Özgür Köroğlu, Ömür Çengelci, Ferdi Keskin, Çağdaş Seçkin, Alisher Khalmatov, Haluk Akgündüz, Yasin Kılınçer, Suat Sevilmiş, Oktay Türinay, Kubilay Avşar, Selvet Bıdırdı, Toygar Karadeniz, Alp Sardağ and Ayşe Genç.

Finally, my life has been constantly fulfilled by love and support of my family. I am extremely grateful to my parents, Şüheda and Halil Akyol, and my wonderful brother, Ayhan Akyol, for their love, caring and sacrifices. Without their endless love and support this thesis would not be possible. This work is dedicated to you.

February 2012

Aydın AKYOL

## TABLE OF CONTENTS

	<u>Page</u>
<b>TABLE OF CONTENTS</b> .....	<b>ix</b>
<b>ABBREVIATIONS</b> .....	<b>xi</b>
<b>LIST OF TABLES</b> .....	<b>xiii</b>
<b>LIST OF FIGURES</b> .....	<b>xv</b>
<b>LIST OF SYMBOLS</b> .....	<b>xxi</b>
<b>SUMMARY</b> .....	<b>xxiii</b>
<b>ÖZET</b> .....	<b>xxv</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1. Motivation .....	1
1.2. Problem Definition .....	3
1.2.1. Image formation model .....	4
1.2.2. Super-resolution as an inverse problem .....	6
1.2.3. Consideration of the problem .....	7
1.3. Contribution of the Dissertation .....	9
1.4. Thesis Outline .....	12
<b>2. LITERATURE REVIEW</b> .....	<b>13</b>
2.1. Interpolation Techniques .....	13
2.2. Regularization Techniques .....	17
2.2.1. Cost function perspective .....	17
2.2.2. Statistical perspective .....	20
2.2.2.1. Bayesian approaches .....	21
2.2.2.2. Example-based methods .....	23
2.3. Heuristic Techniques .....	25
2.4. Multi-Frame Super-Resolution Techniques .....	27
2.4.1. Frequency domain methods .....	27
2.4.2. Spatial domain methods .....	28
2.5. Discussion .....	29
<b>3. ROBUST SUPER-RESOLUTION</b> .....	<b>35</b>
3.1. Robust Statistics .....	36
3.1.1. Re-descending M-estimators .....	39
3.2. SRR With Welsch Type Robust Error Norms .....	42
3.2.1. Response functions .....	43
3.2.2. Quadratic norms vs robust norms .....	44
3.2.3. Data synthesis .....	46
3.3. Experiments .....	48
3.4. Conclusion .....	56

<b>4. LEARNING-BASED SUPER-RESOLUTION .....</b>	<b>59</b>
4.1. Definition .....	60
4.1.1. Inference .....	64
4.1.2. Learning .....	66
4.2. GCRF Devoted To SRR.....	68
4.3. Experiments .....	72
4.4. Conclusion .....	77
<b>5. FACE HALLUCINATION.....</b>	<b>81</b>
5.1. Background .....	83
5.1.1. Global image priors .....	85
5.1.2. Learning model parameters.....	86
5.1.3. Inference and reconstruction.....	89
5.2. Combined Model Fitting In Subspaces .....	89
5.2.1. Representation of images.....	90
5.2.2. Reconstruction of shape data .....	91
5.2.3. Reconstruction of texture component .....	94
5.2.4. Combined reconstruction.....	97
5.3. Experimental Results .....	99
5.3.1. Further investigation on algorithmic details .....	105
5.4. Conclusion .....	108
<b>6. CONCLUSION &amp; DISCUSSION .....</b>	<b>111</b>
6.1. Summary And Contributions .....	111
6.2. Future Directions.....	113
<b>REFERENCES.....</b>	<b>115</b>
<b>CURRICULUM VITAE.....</b>	<b>123</b>

## ABBREVIATIONS

<b>AAM</b>	: Active Appearance Model
<b>ANN</b>	: Artificial Neural Network
<b>BBP</b>	: Bayesian Belief Propagation
<b>CCD</b>	: Charge Coupled Device
<b>CRF</b>	: Conditional Random Field
<b>CS</b>	: Compressed Sensing
<b>EFCM</b>	: Edge-Frame Continuity Modeling
<b>FOE</b>	: Field Of Experts
<b>GCRF</b>	: Gaussian Conditional Random Field
<b>GMRF</b>	: Gaussian Markov Random Field
<b>HMM</b>	: Hidden Markov Model
<b>HP</b>	: High Pass
<b>HR</b>	: High Resolution
<b>K-SVD</b>	: K-means Singular Value Decomposition
<b>MAD</b>	: Median Absolute Deviation
<b>ML</b>	: Maximum Likelihood
<b>MLE</b>	: Maximum Likelihood Estimation
<b>MRF</b>	: Markov Random Field
<b>NE</b>	: Neighbor Embedding
<b>NEDI</b>	: New Edge Directed Interpolation
<b>LAZA</b>	: Locally Adaptive Zooming Algorithm
<b>LP</b>	: Low Pass
<b>LR</b>	: Low Resolution
<b>LS</b>	: Least Squares
<b>PCA</b>	: Principal Component Analysis
<b>PDE</b>	: Partial Differential Equation
<b>PDF</b>	: Probability Distribution Function
<b>PET</b>	: Positron Emission Tomography
<b>PSF</b>	: Point Spread Function
<b>RAF</b>	: Resolution Aware Fitting
<b>RMSE</b>	: Root Mean Squared Error
<b>ROI</b>	: Region Of Interest
<b>SAR</b>	: Synthetic Aperture Radar
<b>SIAD</b>	: Smart Interpolation by Anisotropic Diffusion
<b>SR</b>	: Super-Resolution
<b>SRR</b>	: Super-Resolution Reconstruction
<b>SVD</b>	: Singular Value Decomposition





## LIST OF TABLES

	<u>Page</u>
<b>Table 3.1</b> : Some popular M-Estimators. . . . .	38
<b>Table 3.2</b> : Some popular Re-descending M-Estimators. . . . .	40



## LIST OF FIGURES

	<u>Page</u>
<b>Figure 1.1</b> :Image formation model. . . . .	4
<b>Figure 2.1</b> :Blocking artifacts along diagonal edges in kernel super-resolution. The original image is decimated by 2x2 and then upsampled by linear interpolation. . . . .	15
<b>Figure 2.2</b> :Structure of the techniques performing interpolation by using explicit classifiers. . . . .	16
<b>Figure 2.3</b> :Graphical model for the MRF model used in [1] to define the posterior distribution of the solution space. $\Phi$ and $\psi$ are referred to compatibility functions and used to model local correlations. .	24
<b>Figure 3.1</b> :The space of all probability distribution on a sample space (denoted with the ellipsoid). (a) Non-parametric statistics: allow almost all possible distributions (restriction is quite limited and this ignorance is represented with an interval) (b) Parametric statistics: define strictly determined distributions (represented with a straight line). (c) Robust Statistics: define a neighborhood of strict parametric statistics by allowing slight fuzziness [2]. . . .	37
<b>Figure 3.2</b> : $\rho(x)$ and $\psi(x)$ functions of some popular M-estimators. . . . .	39
<b>Figure 3.3</b> : $\rho(x)$ and $\psi(x)$ functions of some popular re-descending M-estimators. .	41
<b>Figure 3.4</b> :Comparison of the Lorentzian and Welsch norms. Left belongs to the comparison of $\rho(x)$ functions and right is for the comparison of $\psi(x)$ functions. . . . .	42
<b>Figure 3.5</b> :Derivative features used to impose smoothness in the solution. The feature set consists of 8 filters representing the first 2 order derivatives and intermediate orientations to involve also the diagonal components. . . . .	45
<b>Figure 3.6</b> :Edge and bar features used to extract the HF content while cloning the image details. . . . .	47
<b>Figure 3.7</b> :Performance comparison of the Lorentzian and Welsch type M-esitmators in the SRR scheme given in 3.13. (a) Original HR image (Fireman). (b) LR observation (RMSE=20.58). (c) Bicubic interpolation (RMSE=17.79). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=16.54). (e) Reconstruction with the proposed Welsch type error norm (RMSE=16.13). . . . .	49
<b>Figure 3.8</b> :Performance comparison of the Lorentzian and Welsch type M-esitmators in the SRR scheme given in 3.13. (a) Original HR image (Castle). (b) LR observation (RMSE=18.65). (c) Bicubic interpolation (RMSE=16.56). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=14.68). (e) Reconstruction with the proposed Welsch type error norm (RMSE=14.55). . . . .	50

<b>Figure 3.9</b>	:Performance comparison of the Lorentzian and Welsch type M-estimators in the SRR scheme given in 3.13. (a) Original HR image (Goat). (b) LR observation (RMSE=19.16). (c) Bicubic interpolation (RMSE=16.74). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=15.85). (e) Reconstruction with the proposed Welsch type error norm (RMSE=15.37). . . . .	51
<b>Figure 3.10</b>	:Performance comparison of the Lorentzian and Welsch type M-estimators in the SRR scheme given in 3.13. (a) Original HR image (Airplane). (b) LR observation (RMSE=11.27). (c) Bicubic interpolation (RMSE=9.23). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=5.18). (e) Reconstruction with the proposed Welsch type error norm (RMSE=5.24). . . . .	52
<b>Figure 3.11</b>	:Behavior of the Welsch type evaluation function at different scales. (a) Original HR image. (b) LR observation obtained by 2x2 decimation, PSF blurring with a 5x5 Gaussian kernel $N(0,1)$ and additive white Gaussian noise with $\sigma_n = 15$ . (c-h) Reconstruction results with (3.13) having different scale parameters $c$ between 5 and 25. . . . .	53
<b>Figure 3.12</b>	:The mean-face used as the reference image while cloning the image details. . . . .	54
<b>Figure 3.13</b>	:Face reconstruction results using the proposed reconstruction scheme given in 3.11. (a) shows the original HR face images, (b) includes the LR observations obtained by 2x2 decimation, psf blurring with $N(0,1)$ of size 5x5 and additive white noise with $\sigma_n = 10$ , (c) denotes the results of bicubic interpolation, and (d) consists of the reconstructions obtained by the proposed method. .	55
<b>Figure 3.14</b>	:Reconstruction of the car image by using the proposed method (3.11) and the reference image found from the repository search. (a) Original HR image. (b) The found reference. (c) LR observation obtained by 2x2 decimation, psf blurring with $N(0,1)$ of size 5x5 and additive white noise with $\sigma_n^2 = 10$ (RMSE=19.16). (d) Reconstruction by bicubic interpolation (RMSE=16.28). (e) Reconstruction by the proposed method (RMSE=13.82). . . . .	56
<b>Figure 4.1</b>	:Factor graph representation of the image model given in (4.6). Squares refer to factors, diamonds show the observation variables, and circles are the unknown variables representing the image. . .	62
<b>Figure 4.2</b>	:Weighting function features used in [3] for denoising images. The first three rows show the edge filters and the remaining are bar filters.	64
<b>Figure 4.3</b>	:Derivative features used to impose smoothness in the solution. . .	69
<b>Figure 4.4</b>	:Derivative features used to clone image details through a reference image. . . . .	70
<b>Figure 4.5</b>	:Elongated edge and bar filters used to build the weighting function in the form of a regression. . . . .	70
<b>Figure 4.6</b>	:Reconstruction performance of the GCRF image prior where only the piece-wise smoothness is imposed. (a) Original HR image (Fireman). (b) LR observation (RMSE=20.58). (c) Reconstruction by bicubic interpolation (RMSE=17.79). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=15.98). . . .	73

<b>Figure 4.7</b>	:Reconstruction performance of the GCRF image prior where only the piece-wise smoothness is imposed. (a) Original HR image (Castle). (b) LR observation (RMSE=18.65). (c) Reconstruction by bicubic interpolation (RMSE=16.56). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=13.25). . . .	74
<b>Figure 4.8</b>	:Reconstruction performance of the GCRF image prior where only the piece-wise smoothness is imposed. (a) Original HR image (Airplane). (b) LR observation (RMSE=11.27). (c) Reconstruction by bicubic interpolation (RMSE=9.23). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=4.41). . . . .	75
<b>Figure 4.9</b>	:Reconstruction performance of the GCRF image prior where not only the piece-wise smoothness is considered but also the data cloning constraints are incorporated. (a) Original HR face. (b) Reference HR image which is correctly aligned with the observation. (c) LR observation (RMSE=14.61). (d) Reconstruction by bicubic interpolation (RMSE=12.46). (e) Reconstruction by inferring from the posterior given in (4.23) (RMSE=10.77). . . . .	76
<b>Figure 4.10</b>	:Reconstruction quality degrades when the data cloning constraints are neglected. (a) Original HR face. (b) LR observation (RMSE=14.62). (c) Reconstruction by bicubic interpolation (RMSE=12.46). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=10.71). . . . .	77
<b>Figure 4.11</b>	:Reconstruction performance of the GCRF image prior where not only the piece-wise smoothness is considered but also the data cloning constraints are incorporated. (a) Original HR image. (b) Reference HR image which could be aligned with the observation roughly. (c) LR observation (RMSE=19.22). (d) Reconstruction by bicubic interpolation (RMSE=16.31). (e) Reconstruction by inferring from the posterior given in (4.23) (RMSE=13.39). . . .	78
<b>Figure 4.12</b>	:Reconstruction quality degrades when the data cloning constraints are neglected. (a) Original HR image. (b) LR observation (RMSE=19.20). (c) Reconstruction by bicubic interpolation (RMSE=16.27). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=13.32). . . . .	79
<b>Figure 4.13</b>	:Reconstruction performance under different noise levels. . . . .	80
<b>Figure 5.1</b>	:Comparison of image prior models in terms of representational power and computational complexity. Continuous lines show the corresponding behavior for different topologies, and the dash line is the target behavior of this work. The topmost row shows roughly the topologies used. The symbol $\rho$ refers to the distribution models for local image regions, and $\psi$ is the transition model between these local regions. . . . .	82

<b>Figure 5.2</b>	:Reconstruction results with different image prior models. a) HR ground-truth, b) LR observation, c) Result with spatially invariant (homogeneous) local model [4], d) Result with spatially variant (heterogeneous) local model [5], e) Result with global model [6]. Note that only the region of interest (ROI) is reconstructed and the rest is re-sampled from the noise-free LR observation. . . . .	84
<b>Figure 5.3</b>	:Two step processing during in image warping. . . . .	87
<b>Figure 5.4</b>	:Illustration of the biased estimate, occurring when the degradation, $H$ , is not adapted to the alignment. In (a) HR and LR images are aligned individually as $G_H$ and $G_L$ (the spatial mappings, $T_H$ and $T_L$ , were designed as simple global translations; +2 pixels horizontal and +3 pixels vertical). $G_L$ is the ground-truth in comparisons with (b) and (c). In (b) the degradation, $H$ , is used in order to obtain the LR form of $G_H$ by $G'_L = HG_H$ . Observe that the resulting $G'_L$ is different from the expected $G_L$ of (a). In (c) the same operation is repeated with the corrected version of the degradation, $H_G$ , by; $G''_L = H_G G_H$ . Now, the result, $G''_L$ , is the same as $G_L$ . Note that the error in $G'_L$ will be greater when complex spatial mappings are in question. . . . .	88
<b>Figure 5.5</b>	:Illustration of the number of landmarks in the detail level of the modeling. More landmarks create more local regions which could be treated individually. In (a) the lip part can be represented with only two local models while in (b) many more models can be employed for the same region. . . . .	92
<b>Figure 5.6</b>	:Image warping causes changes in both intensity values and locations of pixels. Therefore, the image deformation $H$ , providing a linear mapping between the pixels at different resolutions, should be adapted to this change. Otherwise, when $H$ is used with the aligned textures $G_L$ and $G_H$ in (5.22), the reconstruction in (5.23) results in biased estimates. In column (b) the LR textures $G'_L$ have been obtained by using the image deformation operator $H$ as: $G'_L = HG_H$ . These textures are different from the references in column (a). In column (c), the corrected version of the deformation $H_G$ is used to build the LR texture $G''_L$ by: $G''_L = H_G G_H$ . The resulting textures are almost the same with the textures in column (a). . . . .	96
<b>Figure 5.7</b>	:Summary of the processing followed by the proposed SR approach.	99
<b>Figure 5.8</b>	:Shape-free texture syntheses. . . . .	102
<b>Figure 5.9</b>	:Qualitative results for the reconstructions from the input LR images shown in the left-most column. Note that for all columns the background has been obtained from the linear interpolation of the noise-free LR image to make the ROI more detectable. . . . .	103
<b>Figure 5.10</b>	:Root Mean Square (RMS) errors in texture subspace representations (it is also statistical summarized on the left through the box-plot representation). Note that 100 test samples have been obtained by following a leave-one-out strategy. . . . .	104

<b>Figure 5.11</b> :RMS errors in shape subspace representations (it is also statistical summarized on the left through the box-plot representation). Note that 100 test samples have been obtained by following a leave-one-out strategy. . . . .	104
<b>Figure 5.12</b> :Effect of using the corrected form of the deformation operator on texture synthesis performance. (a) The real HR texture (b) Synthesized HR texture by using $H_G$ in texture reconstruction (c) HR texture synthesis by using the image $H$ in texture reconstruction. . . . .	106
<b>Figure 5.13</b> :RMSEs for the texture subspace representations, found by using the corrected deformation $H_G$ in Procedure 3, compared with the error rates of the texture estimations, obtained by using the image deformation $H$ in Procedure 3. . . . .	107
<b>Figure 5.14</b> :Effect of incorporation of shape into the reconstruction. Dash-line shows the error rates of the subspace representations of the texture component, which is found by employing (5.29) and neglecting shape information. Continuous line show the error rates of the texture reconstructions obtained from the combined solution proposed in Procedure 3. . . . .	107
<b>Figure 5.15</b> :Reconstruction results of naturally degraded LR observations. $e_s$ and $e_t$ are the RMSEs in the subspace representations of the shape and texture components, respectively. Note also that for last 3 columns the background has been obtained from the linear interpolation of the noise-free LR image to make the ROI more detectable. . . . .	109





## LIST OF SYMBOLS

$I$	: Image to be estimated (used when talking on generic inverse problems)
$O$	: Image observed (used when talking on generic inverse problems)
$I_L$	: Low Resolution (LR) image
$I_H$	: Low Resolution (HR) image
$\sigma$	: Variance
$\Sigma$	: Covariance matrix
$\mu$	: Mean
$l$	: Size of one side of a squared image in LR
$h$	: Size of one side of a squared image in HR
$\Omega(x)$	: Irradiance function
$\xi(x)$	: Point spread function
$\eta(x)$	: Noise function
$o(x)$	: Optical effects
$a(x)$	: Spatial integration in sensor area
$M$	: Magnification factor
$\zeta(x)$	: Correspondence function between the LR and HR image planes
$\text{int}[]$	: Quantization operator
$z, p, q$	: Pixels representing the coordinates in 2D, such as $p = (p_x, p_y)$
$D$	: Decimation operator
$B$	: Blurring operator
$n$	: Noise operator
$H$	: Degradation operator
$p(x)$	: Probability distribution function
$\alpha, \beta, \gamma$	: Mixing weights
$\Gamma$	: Image feature convolution kernel
$\mathbf{\Gamma}$	: Image feature operator referring the convolution : with the kernel $\Gamma$ at all pixels uniformly
$\nu$	: Image feature convolution kernel (for weight function)
$\varphi$	: Interpolation kernel
$\epsilon$	: Error rate
$\rho(x)$	: Evaluation function (generic)
$\mathbf{D}$	: Over-complete dictionary
$g$	: Sparse code
$Z$	: Normalization constant
$U$	: Energy function
$\Theta$	: Parameter set
$\theta$	: Function parameters
$\kappa$	: Function parameter
$\Psi$	: Neighbor compatibility function
$\Phi$	: Observation compatibility function

$\mathbf{X}$	: CFT coefficients of an image
$\mathbf{Y}$	: DFT coefficients of an image
$\Lambda$	: Transform operator(matrix) between the CFT and DFT coefficients
$\nabla$	: Gradient
$g(x)$	: Diffusion function
$ n_s $	: Spatial neighborhood of a pixel $s$
$\psi(x)$	: Influence function
$f(x)$	: Distribution function
$c$	: Scale parameter
$J(x)$	: Evaluation function
$r(x,y O)$	: Response estimator at point (x,y) given the observation
$R$	: Concatenation of all response estimators ( $r_i$ 's)
$S$	: Reference image
$F$	: Concatenation of all image feature operators ( $\mathbf{F}$ )
$\mathbb{E}$	: Energy function
$w(x)$	: Weight function
$W$	: Weighting operator
$T$	: A sample image from the training set
$LL(x)$	: Log-likelihood function
<b>ExpVal[]</b>	: Expected value
$C(x)$	: Cost function
$L(x,y)$	: Loss function
$\uparrow$	: Up-sampling operation
$G$	: Texture component
$X$	: Shape component
$t$	: Subspace representation of the texture component
$s$	: Subspace representation of the shape component
$M$	: Subspace transform operator for the texture component
$N$	: Subspace transform operator for the shape component
$e$	: Representational gap in texture subspace representation
$\varepsilon$	: Representational gap in shape subspace representation
$T(X)$	: Spatial mapping of the shape data $X$
$W(I, T)$	: Texture warping on image $I$ based on the spatial mapping $T$
$v$	: Total observation gap
$d(x)$	: Displacement function
$Q$	: Joint subspace transform operator of the transformed image components
$\mathbf{P}$	: Sub-block in $Q$ corresponding to the transformation of the shape component
$\mathbf{R}$	: Sub-block in $Q$ corresponding to the transformation of the texture component
$\omega$	: Scaling factor
$\mathbf{c}$	: Joint representation of shape and texture subspace representations
$\mathbf{t}$	: Delaunay triangle
$b_{t_i}(X_t(x,y))$	: $i$ th barycentric coordinate of the $t$ th triangle on the shape data $X$
$a$	: Subspace coding
$\eta$	: Representational gap (within the transformation via $Q$ )
$\Upsilon$	: Constant linear relationship
<b>res</b>	: Residual
$Z(k)$	: $k$ th region
$\tau$	: Regression coefficients

# **EFFICIENT TECHNIQUES FOR THE SINGLE-FRAME SUPER-RESOLUTION RECONSTRUCTION OF INTENSITY IMAGES**

## **SUMMARY**

In many cases, the imaging sensors have outputs in poor resolution, which is not sufficient for accurate machine/human perception. At that point, hardware solutions remain incapable of enhancing the resolution at desired levels, and Super-Resolution Reconstruction (SRR) techniques are referred.

SRR is an ill-posed inverse problem and requires the estimation of large-scale unknowns. The exact solution is approximated by regularizing the solution space through additional constraints. A typical SRR method consists of three main components: the constraints to be imposed, the optimization technique used to maximize the objective function under these constraints and the trusted data source to be used for extrapolation. Constraints and the data source are mainly related with the accuracy of the resulting estimator, while the optimization technique determines the computational complexity of the method. It is known that the natural image space has a heterogeneous nature and requires adaptive treatment of local image regions. However, growing adaptation means not only an increase in complexity and number of the constraints but also folding in the difficulty of the optimization. Despite this conflicting relation, almost all applications desire an SRR method, which is both computationally simple and highly accurate. In addition to quality and complexity, the needs and the available resources (adequate data for learning, time constraints and the generality of the imaging space) affect the practicality of a solution.

This thesis provides efficient single-frame SRR techniques that are computationally simple and provide reconstructions of high-quality for varying scenarios. First, we consider the scenario where the imposed constraints are adjusted manually; hence, no learning is needed. An iterative reconstruction scheme that benefits from robust statistics is proposed. The Welsch norm, having strict edge-stopping utility and computational conveniences, is used for the imposed constraints to exhibit heterogeneous behavior. Later, we consider the case where the constraints are learned from data rather than being set manually. We propose using an enhanced image prior model based on the Gaussian Conditional Random Field (GCRF). The selected GCRF modeling scheme provides significant computational advantages, and the reconstruction can be obtained analytically. In another case we address SRR for the constrained image domains, where the training and test data are strictly correlated. An efficient method is built in subspaces by employing generative models and utilizing shape and texture components together. The main idea here is that the image details can be synthesized by global modeling of accurately aligned local image regions. In order to achieve sufficient accuracy in alignment, shape reconstruction has been considered as an individual problem and solved together with texture reconstruction in a coordinated manner. Meanwhile, the statistical dependency between shape and texture components is also considered. Moreover, different from traditional

model-based SRR methods, we employ a corrected form of the degradation operator with the aligned images. It is shown that when the degradation operator is used with the aligned texture components as is, the least-squares solution results in biased reconstructions. To overcome this problem, we reflect the same processing, performed in alignment, onto the degradation operator, and use this corrected version in texture reconstruction.

Throughout the thesis, globally consistent structures are utilized as the data source for extrapolation. Thus, the difficulties with the use of local image models and insufficient dictionary schemes are avoided.

## TEK İMGEDEN SÜPER-ÇÖZÜNÜRLÜKLÜ GERİ-ÇATMA AMACIYLA GELİŞTİRİLMİŞ ETKİN YÖNTEMLER

### ÖZET

Kamera duyarga yapılarının oluşturdukları imgeler, pek çok durumda hem imge analizine gerek duyan uygulamalar için hem de insan algılaması için yeterli çözünürlükte değildir. Bu noktada, çözünürlüğün artırılması için donanım ile üretilecek çözümler de yetersiz kalır ve Süper-Çözünürlüklü Geri-Çatma (SÇG) tekniklerinden faydalanılır.

SÇG eksik koşullandırılmış ters bir problemdir ve büyük belirsizlik oranlarının kestirimini gerektirir. Bu amaçla, imge modelleri ile ek kısıtlamalar yaratılıp çözüm uzayının mümkün olduğunca düzenlenmesi yoluna gidilir. Tipik bir SÇG çözümünün 3 temel bileşenden oluştuğu söylenebilir: uygulanacak kısıtlar, bu kısıtlar ile beraber oluşacak hedef fonksiyonun eniyilenmesinde kullanılacak teknik, ve dışdeğerleme için faydalanılacak veri kaynağı. Kısıtlar ve veri kaynağı oluşacak kestiricinin doğruluğu ile daha çok ilgili iken, eniyilemede kullanılacak teknikler de hesaplamadaki basitlik ile büyük oranda ilgilidir. Doğal imge uzayı çoktörel bir yapıya sahiptir ve bu nedenle yerel imge alanları için ayrı ayrı uyarlanabilen işlemlerin kullanımına gereksinim duyar. Ancak, uyarlanmadaki artış hem uygulanacak kısıtların karmaşıklıklarının artması hem de eniyilemenin kat ve kat zorlaşması anlamına gelir. Aradaki bu çelişik ilişkiye rağmen, hemen hemen tüm uygulamalar geri-çatma kalitesi yüksek ve hesaplama maliyeti düşük SÇG yöntemlerini arzular. Kalite ve hesaplama maliyetine ek olarak, ihtiyaçlar ve eldeki olanaklar da (eğitim için yeterince verinin olması, zaman kısıtları ve üzerinde çalışılan imge uzayının büyüklüğü gibi) çözümün pratikliğini etkilerler.

Bu tez kapsamında, farklı durumlarda düşük maliyetle yüksek kalitede geri-çatma sağlayabilecek verimli SÇG teknikleri oluşturulmuştur. Önce, kullanılacak kısıtların baştan ayarlanabildiği ve böylece eğitime gerek kalmayan durumlar için, gürbüz istatistik fonksiyonları kullanılarak yinelemeli bir çözüm oluşturulmuştur. Uygulanan kısıtların çoktörel bir davranış sergilemesi amacıyla, etkin bir ayırıcılığa ve hesaplama kolaylıklarına sahip olan Welsch tipi fonksiyonun kullanılması önerilmiştir. Daha sonra, kullanılacak kısıtların baştan ayarlanması yerine, eldeki veriden öğrenilmesi şeklinde bir çözüm oluşturulmuştur. Önerilen bu çözümde, adaptasyonun artırılması amacıyla, geliştirilmiş Koşullu Gauss Tipli Markov Rastgele Alanı temelli bir imge modeli oluşturulmuştur. Seçilen imge modelinin hesaplama avantajları sayesinde, analitik bir geri-çatma ifadesi ile çözüme gidilebilmiştir. Ele alınan diğer bir durumda da, kısıtların öğrenilmesinde kullanılan veriler ile test verisi arasında daha sıkı bir ilişki mevcuttur. Örneğin, kısıtlanmış imge uzaylarında (sadece yüz imgelerinden oluşan uzay gibi) geri-çatma ihtiyacı bu yapıda bir durumdur. İşte bu türden kısıtlanmış imge uzayları için, alt-uzayda tanımlanmış üretken modellere dayanan ve hem şekil hem de doku bileşenlerini kullanan verimli bir yöntem sunulmuştur. Buradaki temel fikir, imge detaylarının doğru hizalanmış yerel imge alanlarının

bütünsel modellenmesi ile sentezlenebileceğidir. Hizalamada yeterince doğruluğa erişebilmek amacıyla, şekil bilgisindeki geri-çatma ayrı bir problem olarak ele alınmış ve doku bileşeninin geri-çatma problemi ile beraber koordineli çözülmüştür. Bu arada, şekil ve doku bileşenleri arasındaki ilinti de çözüme katılmıştır. Ayrıca, geleneksel model-tabanlı yaklaşımlardan farklı olarak, deformasyon operatörünün hizalanmış imgeler için özel olarak ayarlanmış hali çözümde kullanılmıştır. Deformasyon operatörünün hiç düzeltme yapılmadan hizalanmış imgeler ile kullanımı söz konusu olduğunda, en-küçük kareler çözümü ile elde edilen geri-çatmanın yanlış olduğu deneylerle gösterilmiştir. İşte, bu problemin üstesinden gelmek amacıyla, hizalama sırasında yapılan işlemler deformasyon operatörüne de uygulanmış ve doku bileşeninin geri-çatılmasında bu yeni sürüm kullanılmıştır.

Tez boyunca, dışdeğerleme için kullanılacak veri kaynağı seçiminde, bütünsel sürekliliğe sahip daha gerçekçi yapıların kullanılması önerilmiş, böylece yerel modellerin ve örnek-sözlüklerinin kullanımlarındaki zorluklardan kaçınılmıştır.

# 1. INTRODUCTION

## 1.1 Motivation

Both the human and machine perception are based on image analysis where the meaningful information is extracted from images to characterize them quantitatively or qualitatively. There are many different techniques used in automatic analysis of images such as: object recognition, segmentation, tracking, detection, pose estimation. These techniques have continuously expanding applications throughout all areas of science and industry, including:

- Security and defense: target detection and missile guidance, unmanned vehicles, intruder detection and border observation, biometric security, face recognition, license plate recognition, etc.
- Medicine: diagnostics (e.g. detecting cancer in an MRI scan), microscopy (such as counting the germs on a swab), etc.
- Industrial machine vision and robotics: industrial automation (e.g. counting items on a factory conveyor belt), inspection (e.g. determining cracks if a metal weld has), material analysis (e.g. determining the mineral content of a rock sample), topographical modeling, identification of outliers, sensing, cybernetics, etc.
- Entertainment, Internet, and media: 2D/3D games, human-computer interaction, data coding, data compression, data conversion, photography, etc.
- Astronomy: observation, event detection (e.g. detection of solar features and sunspots), recognition (e.g. calculating the size of a planet and characterization of galaxies), etc.

One of the key factors for success in image analysis is the amount of available informative data, which is determined by the resolution concept in imaging. As the

resolution increases, the quality of the analysis gets higher. Moreover, the demand for further quality would never end as the applications get more sophisticated and talented.

Opposed to the need for an excessive amount of visual information, it is known that the optics of an imaging system limits the amount of information received by the imaging device [7]. These imaging systems yield aliased and under-sampled images since their detector arrays are not sufficiently dense. At that point, increasing the number of pixels per unit area or increasing the chip size could be thought as viable solutions, but unfortunately both have limits [8]. For instance, as the pixel size decreases, the amount of light available decreases and it causes shot-noise severely degrading the image quality. Similarly, increasing the chip size would not only be expensive but also lead to an increase in capacitance, which results in slowing down the charge transfer rate.

Today's excessive demand for higher resolution images and saturation in imaging device technology require better and faster Super-Resolution Reconstruction (SRR) techniques, which are defined as intelligent techniques transcending the limitations of imaging systems, much more than anytime. Although there has been strong research on the problem during the past three decades, we are still far from the solution valid for any real-world scenario. As explained in Chapter 2, the main difficulty in the problem is caused by mainly two factors: the ill-posedness of the reconstruction and the large-scale unknowns.

The exact solution for the ill-posed problems can only be approximated via regularization, either deterministically or statistically. However, the natural image space does not show any particular regularity to be modeled, except being piece-wise smooth. This dispersed nature of the image space requires individual treatment of local image regions, and this means more complex models are needed. In general model complexity involves a trade-off between simplicity and accuracy of the model. While added complexity usually improves the realism of a model, it can make the model difficult to analyze and pose computational problems. Occam's razor [9], which is a principle particularly relevant to modeling, states that among models with roughly equal predictive power, the simplest one is the most desirable. So, we



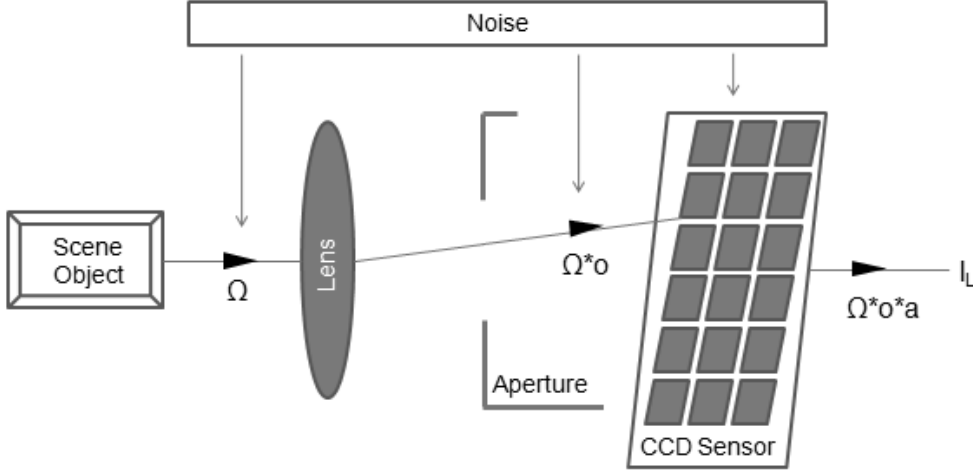
should tend toward simpler models until we can trade some simplicity for increased representational power.

In this thesis, we focus on efficient methods which maximize the accurate extrapolation for image quality, while keeping the computational cost at acceptable levels for real-world practical applications. Considering these two conflicting goals, we design solutions based on the following principles:

- In mathematical programming literature, the most efficient solutions can be obtained via quadratic programming. So, to keep the computational complexity low enough, we should always tend to use quadratic objective functions.
- As will be explained in Chapter 2, the general tendency in image modeling is to find out regularities for local image regions since the variety soars as the size of the images increases. However, these local models generally suffer from global discontinuity artifacts in addition to the computational burden. So, we should employ data sources, providing globally continuous and realistic textural data (e.g. a repository image having structurally and semantically similar content) to extrapolate.

## **1.2 Problem Definition**

In a generic sense, image super-resolution is considered as a reconstruction problem under the assumption of a linear relation, called observation model (known also as forward model or formation model), between the LR and HR images. The assumptions made initially on this model and the other components of the problem setup highly affect the solution strategies to be followed. Due to this variety, it is hard to give a single definition for the generic SRR problem. So, the problem is described together with the assumptions made initially. We give our assumed observation model in Section 1.2.1 and the corresponding inverse reconstruction model in Section 1.2.2. Also, in Section 1.2.3 we provide the list of decision points characterizing an SRR problem setup and define ours based on the selections from this list.



**Figure 1.1:** Image formation model.

### 1.2.1 Image formation model

In this section, the generic linear relationship between the discrete Low Resolution (LR) image  $I_L$ , and the original discrete High Resolution (HR) image  $I_H$  is defined. Here,  $I_L$  and  $I_H$  denote the lexicographical ordering of the images and have the sizes of  $[l^2 \times 1]$  and  $[h^2 \times 1]$ , respectively. Within the scope of this dissertation, we restrict ourself to only intensity images and neglect the advantages and disadvantages of other imaging spaces such as X-ray, SAR, PET, ultrasound.

The modeling starts with giving the relationship in continuous domain. The continuous image formation can be visualized as in Fig. 1.1, where the LR image  $I_L$  is formed by the convolution of the irradiance  $\Omega(x)$  with the camera Point Spread Function (PSF)  $\xi(x)$ , and the additive environmental noise function  $\eta(x)$ .

The PSF is modeled as the convolution of the optical effects  $o$  (caused by the lens and the finite aperture) and the spatial integration performed on the sensor area  $a$  (assumed square and uniformly sensitive to light) as  $\xi(x) = (o * a)(x)$  [10]. Although PSF is a very complex function which depends upon a large number of parameters, in practice a simple parametric form is assumed for  $\xi(x)$ , more often than not, it is Gaussian,  $N(0, \sigma_\xi^2)$  [11]. Moreover in Super-Resolution we want to estimate  $\Omega$  on a denser grid to enhance the resolution by the linear magnification factor  $\mathbb{M} = \frac{h}{l}$ . Considering that  $\zeta(z) = \frac{z}{\mathbb{M}}$  is used for the correspondence between the LR image plane and the super-resolved image plane, the continuous form of the observation model can be

formulated as

$$I_L[q] = \int \Omega(\zeta(z)) \cdot \xi(\zeta(z) - q) \left| \frac{\delta \zeta}{\delta z} \right| dz + \eta(q), \quad (1.1)$$

where  $z = (z_x, z_y) \in \mathbb{R}^2$  refers to the points in continuous image domain,  $q = (q_x, q_y) \in \mathbb{Z}^2$  refers to the points in the discrete observation  $I_L$ , and  $\eta(q)$  represents the total additive noise at point  $q$  [10]. Note also that the integral is defined on the super-resolved image plane and  $\left| \frac{\delta \zeta}{\delta z} \right|$  is the determinant of the Jacobian. Here,  $\Omega(\zeta(z))$  corresponds to the irradiance, that would have reached the image plane of the camera under the pinhole model and transformed onto the super-resolution image plane. Since the source scene does not change,  $\Omega(\zeta(z))$  can be considered as  $\Omega(z)$ . Including this update the continuous formation model can be rewritten as follows

$$I_L[q] = \frac{1}{\mathbb{M}^2} \cdot \int \Omega(z) \xi(\zeta(z) - q) dz + \eta(q). \quad (1.2)$$

To proceed for the formal SRR definition, we need to specify the continuous function  $\Omega(x)$  with a discrete image  $I_H$ . In the simplest case  $I_H$  represents the piecewise constant function;  $\xi(z) = I_H[p]$  for all  $z \in (p_x - 0.5, p_x + 0.5] \times (p_y - 0.5, p_y + 0.5]$ , where  $p = (p_x, p_y) \in \mathbb{Z}^2$  refers to discrete points in the HR image,  $I_H$ . Now we can re-organize the image formation model (1.2) by using the discrete representation of irradiance function as

$$I_L[q] = \sum_p I_H[p] \cdot \frac{1}{\mathbb{M}^2} \cdot \int \xi(\zeta(z) - q) dz + \eta(q). \quad (1.3)$$

Images are always intensity discretized (typically to 8-bit values in the range of 0-255 gray levels). Therefore, there will always be some perturbations in the observation, even when the additive environmental noise  $\eta(x)$  does not exist [10]. Supposing that  $\mathbf{int}[\cdot]$  denotes the quantization operator, then the noiseless measurement would actually be

$$I_L[q] = \mathbf{int} \left[ \sum_p I_H[p] \cdot \frac{1}{\mathbb{M}^2} \cdot \int \xi(\zeta(z) - q) dz \right]. \quad (1.4)$$

However, it is common to denote this error as part of the additive Gaussian noise  $\eta(x)$  as in (1.3). In fact, while other distributions for noise are possible, the Gaussian distribution is still usually a good model due to the Central Limit Theorem. There are

multiple sources of noise (e.g. read-out in CCD, atmospheric turbulence, transmission, quantization as mentioned above, sensor heat) and their sum can be approached well with a Gaussian distribution.

In practice, to make the problem more tractable, both the observation model (1.3) and the reconstruction model are often represented in discrete space. The operators in (1.3) can be approximated in discrete domain as matrices

$$\begin{aligned} \text{Decimation} & : \zeta(x) \longrightarrow D, \\ \text{PSF Blurring} & : \xi(x) \longrightarrow B, \\ \text{Noise Function} & : \eta(x) \longrightarrow n, \end{aligned} \tag{1.5}$$

and when they are substituted on the continuous image formation model (1.3), we obtain the completely discretized forward model as

$$I_L = DBI_H + n. \tag{1.6}$$

Assuming that the observation  $I_L$  is of size  $[l^2 \times 1]$  and the HR super-resolved image  $I_H$  is of size  $[h^2 \times 1] = [\mathbb{M}^2 l^2 \times 1]$ , then the other terms of equation (1.6) would be;  $D \in \mathbb{R}^{l^2 \times \mathbb{M}^2 l^2}$ ,  $B \in \mathbb{R}^{\mathbb{M}^2 l^2 \times \mathbb{M}^2 l^2}$  and  $n \in \mathbb{R}^{l^2 \times 1}$ . Note also that it is common to denote the blurring and decimation operators together within a single deformation operator  $H = DB$  as

$$I_L = HI_H + n, \tag{1.7}$$

where  $H \in \mathbb{R}^{l^2 \times \mathbb{M}^2 l^2}$ .

### 1.2.2 Super-resolution as an inverse problem

Given the observation and the forward model, the goal of SRR is to estimate the  $I_H$ , which is in higher resolution than the observation. Theoretically, this corresponds to the inverse of the image formation model and can be represented as

$$\hat{I}_H \cong \arg \min_{I_H} \|I_L - HI_H\|_2^2, \tag{1.8}$$

where the operator  $\|\cdot\|_2^2$  refers to the square of the L2-norm. However, since lots of ambiguities are included and the image degradation operator is singular, the inverse of the forward model can not be found analytically. It is easy to see from (1.3) and (1.6) that the main sources of these ambiguities are

- PSF blurring: PSF is assumed to be as a smoothing operator which realizes a uniform Low Pass (LP) filter,
- Decimation: The observation model, either in continuous or discrete form, reveals that number of unknowns is much more than the number of measurements (e.g. for 2x2 decimation 75% and for 4x4 decimation 93.75% of the data to be synthesized),
- Noise: Even the simplest form of the noise term is enough to make the problem badly conditioned,
- Quantization: The standard rounding operator, **int[]**, replaces a real number with the nearest integer. It was shown [10] that the volume of the set of solutions of (1.3) grows asymptotically with the number of pixels on the HR grid.

In order to have a mathematical answer, a typical inverse problem should satisfy the solution existence, uniqueness and stability. However, none of these conditions is satisfied in the SRR case. Though there are thorough studies on the conditioning analysis of the SRR problem, such as [10] and [12], it is apparent from the above list of ambiguities that the inverse problem is not tractable.

However, it is possible to regularize the inversion process and approximate the true solution by imposing additional constraints. In deterministic and statistical perspectives, the generic regularization framework can be given as in (1.9) and (1.10), respectively

$$\hat{I}_H \cong \arg \min_{I_H} \|I_L - HI_H\|_2^2 + \lambda \|\mathbf{\Gamma} I_H\|_2^2, \quad (1.9)$$

$$\hat{I}_H \cong \arg \max_{I_H} p(I_L|I_H)p(I_H), \quad (1.10)$$

where  $\lambda \|\mathbf{\Gamma} I_H\|_2^2$  and  $p(I_H)$  refer to the regularization terms. More details on these expressions are provided in Section 2.2.1 and Section 2.2.2.

### 1.2.3 Consideration of the problem

The regularization term (corresponds to *a priori* information in statistical perspective) is designed based on the assumed problem setup. The following list of characteristics shape a problem setup.

- Number of observations: There are two considerations; some researchers do not accept such a diversification and define the SRR for only the case having multiple observations [13, 14, 15, 16, 17, 18], while some others believe that the single frame SRR is the main problem and the multi-frame SRR is a special case of it.

Having multiple observations means having more information about the solution, and this additional information can be exploited to regularize the solution more. To be informative an observation should be shifted with sub-pixel precision. Though it is not always easy, when the shutter speed and the camera calibrations are adjusted appropriately (either by capturing the same scene with the same camera at different times or capturing the scene with different cameras having similar camera parameters), this kind of observations can be obtained. In addition to the difficulties in image capturing, the use of multiple frames requires an additional pre-processing, called registration. But, registration is as intractable as the reconstruction, so the difficulty of the problem is doubled.

Some researchers prefer working on single-frame SRR problem by renouncing the aliasing information. However, this renunciation would compel them to look for additional data sources to extrapolate. A detailed investigation of the methods proposing alternative designs for the reference data source is provided in Chapter 2.

- The imaging space under consideration: The generality of the imaging space significantly affects the solution strategy. For instance, when the natural image space is considered, the estimation of the unknown pixels in HR turns to almost random guessing since natural images show no particular regularity. On the other hand, when the image space constrained to a specific domain, common characteristics of the domain images can be incorporated into the solution.
- Knowledge about the imaging environment: Most SRR methods assume that the degradation parameters are already known. Even if we do not know, we can reach acceptable estimations by using generic models or may approximate the true parameters empirically by using simple measurements [10].

In literature, there are also blind-reconstruction methods [15, 19, 10], which consider the degradation parameters as unknowns and jointly estimate them together with the unknown HR image.

- Prior knowledge: In addition to the observation data and the forward model, we may have prior knowledge about the solution and can enforce the intermediate estimates to conform with it. As will be explained in more detail in Chapter 2, regularization techniques exactly define this intent. This prior information can be either: “a set of rules expected to be satisfied [20, 21]”, “a parametric model [22, 23]”, “a distribution function conditioned on the observation or some data source [24, 25]”, “a non-parametric model based on some dictionary [26, 27]” or “any data having clues about the solution (e.g. segmentation map of the solution, class membership information, scene label) [28, 29, 30]”.

During our research we have mainly considered the problem setup having the following features:

- Gray-scale intensity images are considered,
- Single observation exists,
- Image formation model (deformation operator, noise, and the decimation rate) is known,
- Natural image space is considered,
- No pre-set conditions exist.

Though we have mostly used this problem setup, in some chapters we have also employed the slightly deviating versions of it. For instance, in Chapter 5, we consider the image space that is restricted to only frontal face images.

### **1.3 Contribution of the Dissertation**

The main focus of this thesis is the task of super-resolving intensity images given a single observation. As with many other image processing and computer vision problems, the SRR is an ill-posed inverse problem and approached with approximate

models under some artificial constraints. Approximate solutions are shaped with the assumed models and resources available. The reality of the models and the constraints determine the quality of the reconstruction, while the accuracy in their implementation with the available resources identifies the practicality of the method. In Chapter 2, we have given a survey of the past techniques accompanying the similar image formation model, given in (1.7), and identified the basic principles related with reaching the best quality in reconstruction with the cheapest solution. In light of these principles, we have proposed efficient SRR methods needed by different real-world scenarios.

First, we have considered the scenario where a separate learning stage is not possible (due to either generality or limited resources), but some delay can be accepted during online processing. For that purpose, we have proposed an iterative reconstruction scheme where the constraints are set manually and imposed heterogeneously via robust statistics. In fact, the idea is not new and simpler and suboptimal variants of it [15, 31, 32] had been proposed before. But, different from these non-convex structures, we have used the Welsch-type robust error norm which is partially convex and has a more strict edge-stopping utility. Moreover, to reduce the blocking artifacts we have suggested using a wide set of image features consisting of multi-order and multi-oriented derivatives.

Later, we have considered the scenario, where the fastest online reconstruction is required and training is possible. In this offline training, the constraints are learned from similar images rather than set manually. Thus, more realistic constraints could be obtained to regularize the solution better. Different from the past non-linear and non-convex image models [22, 21, 5, 25], we have proposed using a strictly convex quadratic Gaussian distribution function (the Gaussian Conditional Random Field - GCRF) for image modeling. To overcome the drawbacks of the Gaussian type Markov Random Field (MRF), we have employed an enhanced version of it (introduced by Tappen et al. [3]) by first conditioning with the observation via response estimators, and then adding evaluation mechanism through parametric weighting. Thus, without sacrificing the computational advantages we could gain adaptation. Due to the computational advantages of the quadratic structures, the reconstruction scheme could be defined analytically. We have compared our results with other types of fast



analytical approaches, such as kernel interpolation techniques, and seen that our results significantly outperform.

We have also addressed a more specific case where the training data and the online test data are strongly related. This case defines the SRR problem in constrained image domains, such as face, plate, text, cell. The common characteristics of this new image domain is a valuable information and should be definitely utilized in reconstruction. For that purpose, we have developed a quite efficient reconstruction method, based on global image priors. In fact, global topologies are not common in image modeling due to their limited representational power; the general tendency is to use local image models in the form of MRF. However, these generic local models [33, 1, 22] either mostly constitute non-convex structures, which are adaptive but difficult to optimize, or suffer from serious discontinuity artifacts. Different from these locality-based approaches, we have insisted on using global models by increasing their representational power. For that purpose, we have utilized the shape information in addition to the textural data. Shape reconstruction has been considered as an individual problem and solved in a coordinated manner with the texture reconstruction. By modeling all the variables in the reconstruction expression with quadratic Gaussian functions, we have reached a fast analytical reconstruction expression. Moreover, to further decrease the computational cost and to increase the scalability, we have fully transformed this expression onto subspaces via Principal Component Analysis. Hence, the reconstruction has been turned into simple algebraic operations of the small-size matrices.

Another contribution of this thesis is to show how to benefit from globally consistent data sources for extrapolation. For realistic reconstructions some trusted data source, from which reliable image details can be incorporated, should be employed. The general tendency in data source design is to use statistical models either as a collection of codewords (analytical or learning-based) or as a single distribution function. Unfortunately, none of these models is capable enough to synthesize reliable and realistic data since natural image space, even small local regions, is too wide and dispersed to be modeled. Because of insufficiency in representation, results suffer from excessive blurring and discontinuity artifacts. As a remedy to the problems with statistical modeling, we have suggested using a reference HR image which is

structurally and semantically similar to the observation. We have shown that this memory-based technique can successfully incorporate realistic and consistent image details into the result.

## **1.4 Thesis Outline**

The remaining chapters of the thesis are organized as follows:

In Chapter 2, a review of the SRR literature is provided. This investigation is built over the methods assuming a similar image formation model with the one given in (1.7). The review concludes with an evaluation where the expected behaviors of an ideal solution (maximizes the quality and minimizes the computational cost) are listed. Targeting such efficient solutions, throughout the next three chapters the proposed SRR methods are described.

Firstly in Chapter 3, an adaptive reconstruction scheme is presented by utilizing robust statistics. Specifically, the Welsch type re-descending M-estimator is employed for both smoothness and data cloning constraints. Thus an iterative reconstruction scheme has been constructed for the cases where no resources are available for training and some amount of delay is acceptable during online processing. Later in Chapter 4, an approximation to the adaptive treatment of the robust error norms is described by using quadratic expressions. The resulting estimator is defined by employing enhanced GCRF modeling of the image space. In Chapter 5, the problem is considered from a different perspective by taking into account the constrained image spaces. An efficient solution for rigid-object images is described by utilizing both texture and shape components. Meanwhile, some major drawbacks of the traditional applications are revealed.

Chapter 6 is the concluding section. First, the thesis is summarized, and then the contributions are highlighted. Moreover, a discussion is presented to show future directions of the problem.

## 2. LITERATURE REVIEW

In this chapter, a comprehensive review of the techniques for the SRR problem is presented. Although the initial attempts start with the application of the interpolation techniques [34, 35] to image processing, intense interest began after the seminal work of Tsai and Huang [11], in which they used multiple observations to extrapolate. Many techniques have been proposed over the past three decades for both single-frame and multi-frame cases. Our assumed problem setup considers having a single-observation, and in this review chapter we mainly focus on single-frame SRR approaches.

The techniques are discussed in three broad categories: interpolation methods, regularization methods and heuristic approaches. Among these categories, we devote heavy interest to the regularization techniques as parallel to their popularity in literature. There are a couple of reasons making it advantageous against the other options, such as flexibility for modeling a wide range of image formation models, having consistent theoretical foundation and ability to incorporate almost any type of *a priori* information. In addition to the review in this chapter, constrained domain SRR techniques are also investigated separately in Chapter 5.

### 2.1 Interpolation Techniques

Maybe the earliest and most common way of achieving SRR is to use an interpolation kernel. First the observation,  $I_L$ , is located on the dense grid and sparse approximation of the HR image,  $I_H^S$ , is obtained. After that the HR image,  $I_H$ , is approximated by combining instances of the kernel function,  $\varphi$ , at the known discrete samples,  $(x, y)$ , of the dense grid. This linear operation can be shown as

$$I_H[x, y] = \sum_{i=-k}^k \sum_{j=-k}^k \varphi[i, j] I_H^S[x-i, y-j], \quad (2.1)$$

where  $k$  is the one size of square kernel function. Typical choices of the base functions include; linear, cubic, Spline and Lanczos [34, 35]. One common feature of these

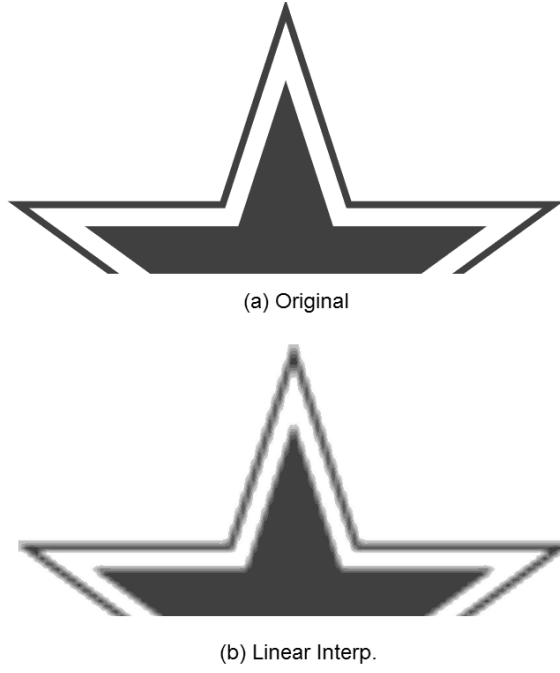
functions is separability, and by means of this feature the reconstruction process can be split in two consecutive steps (horizontal and vertical processing) for reducing the computations.

The implementation of this approach is very efficient because the uniform interpolation kernel can be applied on the input image by using standard matrix operations from linear algebra. This computational convenience makes it popular especially for commercial products. Despite this simplicity in implementation, the results are not always as good as one envisions [36]. Some shortcomings of the kernel-based approaches can be listed as in below.

- Blurring of sharp edges: Kernel filters typically perform very well in smooth areas, but not in edge regions. The reason is apparent; pixels are treated uniformly. To overcome this problem and to capture different characteristics of the image space, adaptive schemes should be employed.
- Blocking artifacts: Blocking artifacts in diagonal edges or lines are caused by the horizontal and vertical orientation of the re-sampling kernels. This limited treatment is unable to recognize diagonal lines, as exemplified in Fig. 2.1. So, to relieve the distortion one should process multiple intermediate orientations at multiple scales.
- Insufficient high-frequency content: High-frequency content corresponds to the image details, and kernel-based methods are not sufficiently powerful to incorporate the necessary details. This extrapolation problem is the most challenging one and requires prior knowledge about the solution. Furthermore, for most cases, the explicit use of some data sources is highly required.

These problems have driven the following and ongoing research for improved super-resolution methods.

In order to overcome the problem of blurring edges, adaptive treatment of the pixels has been proposed rather than filtering uniformly. Explicit functions have been employed to the standard kernel interpolation techniques. The main idea behind this is to use a decision function as a piecewise linear approximation to the conditional mean estimator of the HR image. It is assumed that there are different classes of pixels,

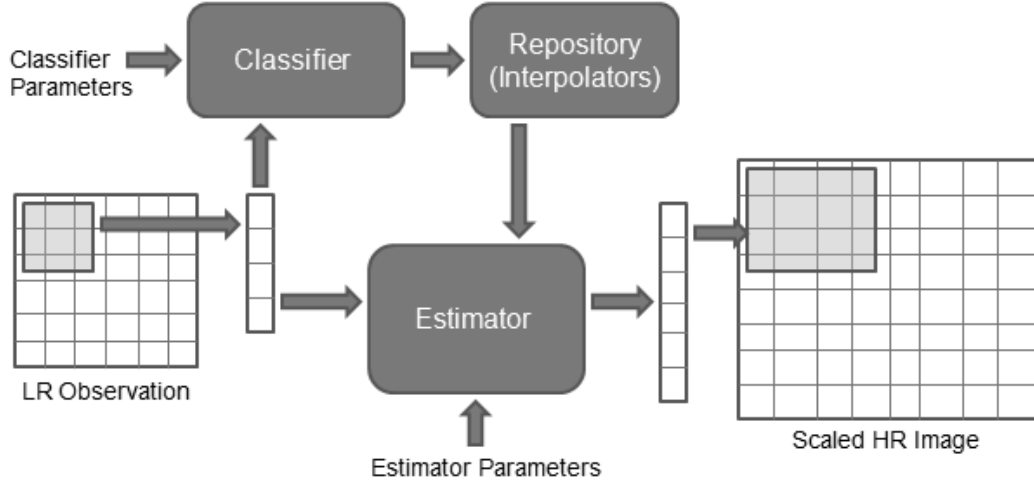


**Figure 2.1:** Blocking artifacts along diagonal edges in kernel super-resolution. The original image is decimated by 2x2 and then upscaled by linear interpolation.

such as the pixels on object edges with different orientations or the pixels in flat areas. Each class requires specific treatment when enlarged and can benefit from a dedicated super-resolution scheme to better preserve edges.

The main flow of these techniques consists of first performing one of many classification schemes from pattern recognition literature and then taking specific actions for each class as shown in Fig. 2.2.

For instance in [37], a decision-tree based classification scheme has been employed. The parameters for the regression tree are found by training on sample images. The pixels are classified into edge and non-edge pixels with different orientations (the consideration of additional orientations relieves also the problem of blocking artifacts). In [38], Atkins et al. has assigned each pixel to multiple classes with different degrees as in mixture of experts strategy. A pixel can for example be classified to be 60% horizontal edge and 40% smooth. For each pixel and class the chance of membership is estimated using a Gaussian distribution. The degree of membership to a class acts as the weight for the result of this linear filter. Similarly, in [39], Self-Organizing-Maps (SOM) has been utilized; first, classification is done by using the SOM method and second, the training of local associative memories, namely the interpolators, takes



**Figure 2.2:** Structure of the techniques performing interpolation by using explicit classifiers.

place for the pixels falling into the corresponding class. In addition, other types of classifiers, such as Support Vector classifiers [40], have been employed as well.

In [41], Artificial Neural Network (ANN) based regressors have been used instead of explicit functions. This non-linear black-box regressor uses a single generalized feed-forward neural network to interpolate. The input of the ANN consists of the pixels in the local window around the source pixel, and the output consists of the pixels constituting the super-resolved image. A multi-layer ANN can achieve recognition of non-linear relations between input and output. Therefore, it is able to preserve edges better and enhance detail more than linear interpolators.

In these classification-based methods, the increased adaptation comes with an increase in the computational load. During the scanning of the HR grid, every pixel/local-region is evaluated by a decision function before the interpolation. Depending on the complexity in the decision process, this evaluation could be quite costly. Moreover, accuracy will always be demanding more classes, and this would make the decision more difficult, as well as the learning. Another drawback with these techniques is related with disregarding the observation model during online processing, and this makes these approaches prone to noise.

## 2.2 Regularization Techniques

The rank-deficient deformation operator  $H$  and the noise,  $n$ , in the observation model, (1.7), make the inverse problem ill-conditioned. In such situations, the set of solutions that adequately fits the data is large and contains many physically unreasonable models. The minimum-norm solution, that is the least-squares (1.8), is unstable, and small changes in the data may lead to large changes in the solution. The answer to these difficulties is found through what is known as regularization. The purpose is to allow inclusion of additional constraints to stabilize the space of possible solutions. A regularization method is often formally defined as an inversion method depending on a single real-valued parameter, which controls the trade-off between solution stability and data fidelity.

There is a wide variety of regularization methods, but an exhaustive treatment is beyond the scope of this chapter, and we provide only a summary of the main ideas. We investigate the regularization methods from two points of view: cost-function perspective and statistical perspective. In the cost function perspective (the algebraic approach), the unknown image is considered to have a deterministic characteristic. While in statistical perspective, both the unknown image and the noise (or any other solution variable if existing), are stationary random variables, and assumed that they have some particular characteristics which can be modeled. For most cases these two perspectives end up with similar optimization problems. For instance, the algebraic least-squares can be interpreted with Maximum Likelihood Estimation (MLE), given the linear measurements corrupted by Gaussian measurement errors.

### 2.2.1 Cost function perspective

These methods are also called generalized least-squares methods. A Least-Squares (LS) problem is an unconstrained optimization problem of such an objective

$$\hat{I}_H = \arg \min_{I_H} \|I_L - HI_H\|_2^2. \quad (2.2)$$

The solution of an LS problem can be reduced to solve a set of linear equations,  $\hat{I}_H = (H^T H)^{-1} H^T I_L$ . There are quite efficient and reliable algorithms for calculating this

analytically. However, in SRR,  $H^T H$  is singular and this means the LS solution is one of the many possible solutions. More information is needed to tune the reconstruction toward a unique solution. Considering the pure LS (2.2) as the simplest unconstrained norm approximation problem, the regularization is a common scalarization method used to solve the bi-criterion problem of

$$\hat{I}_H = \arg \min_{I_H} \rho(\Gamma I_H) \quad \text{subject to} \quad \|I_L - H I_H\|_2^2 \leq \epsilon, \quad (2.3)$$

where  $\rho(x)$  is an evaluation function,  $\epsilon$  is the accepted error threshold (ideally 0). Also,  $\Gamma_i$  is the image feature operator referring to the uniform convolution of the image  $I_H$  with the image feature kernel  $\Gamma_i$  (e.g. derivative kernel) at all pixels;  $\Gamma_i I_H \sim \{\forall(x,y) \in I_H, (\Gamma_i * I_H)(x,y)\}$ . The most common form of the regularization (2.3) is based on Euclidean norm, which results in quadratic programming as in

$$\hat{I}_H = \arg \min_{I_H} \|I_L - H I_H\|_2^2 + \lambda \|\Gamma I_H\|_2^2. \quad (2.4)$$

This is specifically known as the Tikhonov regularization. The use of such quadratic criteria for the regularizer maintains the computational efficiency and leads to an analytical solution

$$\hat{I}_H = [(H^T H) + \lambda(\Gamma^T \Gamma)]^{-1} [H^T I_L]. \quad (2.5)$$

When appropriate PDEs are selected for  $\Gamma$ , the matrix inversion in (2.5) can be easily performed in the frequency domain; since,  $H^T H + \Gamma^T \Gamma$  is block-circulant. Thus, the efficiency of the solution is increased more. Common choices for such kind of  $\Gamma$  are Laplacian [21] and first-order derivatives.

While such linear processing is desirable, it has some disadvantages as well. A common criticism is that the results tend to be overly smooth because the smoothness is imposed uniformly for all pixels. However, it is known that the image space has a non-uniform nature and this homogeneity assumption ignores it. For this reason, the generalization of the Tikhonov approach has been proposed through weights. Despite the gain in adaptation, fast solutions in the frequency domain are no longer possible with these new type of approaches. Another alternative has been proposed through using robust measures instead of quadratic penalty functions. For instance L1-norm [14], the Huber function [42], Cauchy function [22], or Andrew's sine



function [43] have been used as robust evaluation functions. A thorough analysis on their performance is given in Chapter 3. Clearly, with this choice of robust prior, the overall reconstruction algorithm becomes nonlinear and iterative restoration techniques are employed [44]. Another nonlinear reconstruction scheme has been obtained by exploiting the Maximum Entropy in image prior [45] as

$$\hat{I}_H = \arg \min_{I_H} \|I_L - H I_H\|_2^2 + \lambda(I_H \log(I_H)). \quad (2.6)$$

Although the reconstruction with (2.6) results in sharper results than the Tikhonov result, the difficulty again is related with the computational burden of non-linear processing.

Recently a significant amount of interest has been devoted to sparse coding approaches, such as [26, 46, 47]. They are quite similar to the generalized least-squares methods, but this time the optimization is not unconstrained and includes additional sparsity constraints. More particularly, sparse coding methods are based on the statistics of small image patches. An image patch is represented in terms of a linear combination of basis patches selected from an over-complete dictionary  $\mathbf{D}$ . Considering all patches included by the super-resolved image  $I_H$ , this representation refers to  $I_H = \mathbf{D}g$ . In terms of  $g$ , the reconstruction leads to

$$\hat{g} = \arg \min_g \|g\|_p \quad \text{subject to} \quad \|I_L - H\mathbf{D}g\|_2^2 \leq \epsilon, \quad (2.7)$$

where  $0 \leq p < 2$  is the degree of the norm and mostly selected as the L1 norm. Equivalently, (2.7) can be replaced with the Lagrangian form

$$\hat{g} = \arg \min_g \|I_L - H\mathbf{D}g\|_2^2 + \lambda \|g\|_p, \quad (2.8)$$

that replaces the constraint by a penalty. Hence, the regression coefficients  $g$  are forced to be sparse as much as possible. Moreover, some of these methods employ ideas from the compressive sensing theory. Under some strict conditions [48], these methods ensure that linear relationships among high-resolution signals can be precisely recovered from their low-dimensional projections [26, 48].

Compared to other dictionary-based methods, such as Neighbor Embedding methods [49, 27] with a fixed number of neighbors, sparse coding methods adaptively choose

the fewest necessary supports for reconstruction. Thus, over-fitting is avoided and robustness is increased through L1 minimization.

A fundamental consideration in employing sparse coding approaches is the choice of the dictionary  $\mathbf{D}$ . One type of method employs analytic techniques. A mathematical model of the data is formulated, and an analytic construction is developed to efficiently represent the model. This generally leads to dictionaries that are highly structured and have a fast numerical implementation, such as wavelets, curvelets, contourlets, shearlets, complex wavelets and bandelets [46]. Some other approaches employ machine learning techniques to infer the dictionary from a set of examples. In this case, the dictionary is typically represented as an explicit matrix, and a training algorithm is employed to adapt the matrix coefficients to the examples. Algorithms of this type include Generalized PCA [50], the Method of Optimal Directions [51] and the K-SVD [47]. The advantage of this approach is the much finer-tuned dictionary they produce compared to the analytic dictionaries. However, this comes at the expense of generating an unstructured dictionary, which is more costly to apply and to learn.

Except for some of the sparsity-based methods, the methods discussed in this section assume predefined analytical expressions to grasp the complexity of the general image content. One better alternative is to learn these models directly from the data. The following section focuses on this type of method where the regularization is formed based on image examples.

### 2.2.2 Statistical perspective

In the statistical view of the regularization, it is assumed that the noise and the HR image are random variables <sup>1</sup>. Then, the problem is cast as the inference from a posterior distribution. Given the distribution parameters and the observation, the estimate of the HR image will be

$$\hat{I}_H = \arg \max_{I_H} p(I_H | I_L), \quad (2.9)$$

which is known as Maximum-A-posteriori (MAP) estimation. There are two main types of approach; one is directly infer from the posterior distribution as in (2.9), and

---

<sup>1</sup>In fact, the HR image refers to a random field since each pixel is considered individually as a random variable

the other interprets the posterior from the Bayesian perspective and express it as in the form of

$$\hat{I}_H = \arg \max_{I_H} p(I_L|I_H)p(I_H), \quad (2.10)$$

where the posterior is dependent on the likelihood model,  $p(I_L|I_H)$ , and the prior model,  $p(I_H)$ . Note that the denominator  $p(I_L)$  has been neglected since it is considered constant while working on relative probabilities.

As in the cost-function view, the likelihood term captures the fidelity of the estimate to the observation and represented with the noise model. Traditionally, the noise is assumed in the form of additive white Gaussian,  $n \sim N(0, \sigma_n^2)$ , but we consider the more generic case with  $N(0, \Sigma_n)$ , and the likelihood is defined in the matrix form as

$$p(I_L|I_H) \cong \frac{1}{\sqrt{2\pi|\Sigma_n|}} \exp\left(-\frac{1}{2}(I_L - HI_H)\Sigma_n^{-1}(I_L - HI_H)^T\right). \quad (2.11)$$

On the other hand,  $p(I_H)$  captures our prior knowledge about the unknown HR image in the absence of data. This information is used to regularize the solution through a set of constraints. A plethora of image prior models have been proposed, and in the remaining two subsections we investigate some pioneering ones while investigating the statistical SRR methods.

### 2.2.2.1 Bayesian approaches

Bayesian methods allow to naturally incorporate prior information which is based on either some data source or experience based intuition. This information is expressed as a distribution and generally the parametric image models are used. However, it is hard to model whole natural image space with a single distribution due to the huge dimensionality and variety. As a remedy, patch-based image models are referred. First, the local image models are learned, and then the joint image model is derived depending on the assumed topology of these local models. Markov Random Field (MRF) models are the most common tools used for that purpose. As shown by the famous Hammersley-Clifford theorem [9], an MRF model is denoted in the form of Gibbs distribution as

$$p(I_H) = \frac{1}{Z} \exp(-U(I_H, \Theta)), \quad (2.12)$$

where  $Z$  is the normalization constant and  $U$  is a non-negative energy function having the parameter set  $\Theta$ . More details on this relation and the properties of MRFs are given in Chapter 4. These approaches exploit examples to tune the parameters,  $\Theta$ , that control the local priors. In a pioneering work by Zhu and Mumford [52], an MRF has been proposed by considering the following energy function

$$U(I_H, \Theta) = \sum_{i=1}^N \lambda_i \rho(\mathbf{\Gamma}_i^T I_H; \kappa_i), \quad (2.13)$$

which learns on a weighted average of robust measures of smoothness by using different evaluation functions  $\rho(x, \kappa_i)$ , analyzing filters  $\mathbf{\Gamma}_i$ , and weights  $\lambda_i$ . In other words  $\Theta = \theta_1, \dots, \theta_N$ , where  $\theta_i = \kappa_i, \mathbf{\Gamma}_i, \lambda_i$ .

Generally, these MRFs have non-linear and non-convex structures for increased adaptation. Therefore, sophisticated sampling-based algorithms are required for both learning and inference. For instance in [22], Roth and Black have modeled the local potentials with Student-t distribution and performed learning by minimizing the Kulback-Leibler distance between the empirical distribution of the training set and the prior trained. In [52], the parameters have been learned such that the marginals of the prior fit empirical observations, while maximizing the entropy of the distribution function. Moreover, to avoid this computational burden, it is common to use approximate inference by employing gradient ascent methods [53].

On the other hand, the computationally efficient Gaussian MRFs (GMRF) have significant advantages, and inference can be performed analytically guaranteeing the global optimum. However, the basic GMRF [4] suffers from blurring due to excessive smoothing. The conditional random fields have been employed, as in [3], to avoid these difficulties. We provide a detailed introduction for these quadratic Gaussian Conditional Random Field (GCRF) priors in Chapter 4.

Common to all of the above methods is the fact that a parametric energy function is used, and its parameters are tuned by the examples. Also, all these methods call for an involved optimization procedure. Once the regularization expression is ready, it can be deployed for use in the backward model. If the resulting reconstruction expression is non-convex, approximate methods are used without guaranteeing the global optimum. Otherwise, the iterative gradient techniques are employed.

### 2.2.2.2 Example-based methods

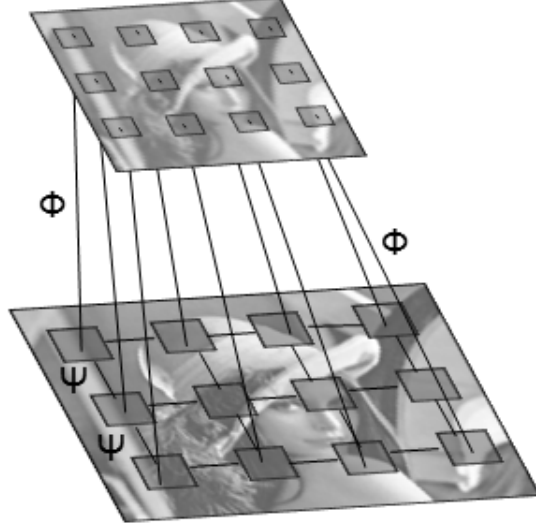
A recently emerging methodology is to use the examples directly within the reconstruction process. Different from the previous parametric approaches, the prior is developed by sampling (samples from the posterior  $p(I_H|I_L)$ ) from other images, and as such a direct way of reconstruction is offered. That is, in these non-parametric (or semi-parametric as in [10]) approaches, the examples are gathered to a database and used explicitly in the on-line reconstruction algorithm.

Though they show slight differences, the main process flow in these approaches starts with pattern matching. Given the LR patch, a database is sought for similar LR examples, and later their corresponding HR pairs are used for the reconstruction. Due to the computational and modeling difficulties (e.g. searching large scale images in big databases), such a process cannot be operated on full size images. Therefore, typically image patches of sizes between 5x5 and 25x25 are used as in [27]. However, there are methods (like [10]), where the above process is operated on a pixel-by-pixel basis. Thus, given the image pairs described above, the LR observation is swept through, and all HR image patches (possibly with overlaps) are extracted [54, 55].

Maybe the most common way of using this idea is Neighbor Embedding (NE), as in [33, 49]. In these works, the LR observation is split into patches without allowing overlaps, and the closest samples are found from the training set via a nearest-neighbor search. Later, the corresponding HR patches of the found LR database patches are aligned sequentially. Since no overlapping in both LR and HR images are considered, the solution is straightforward. However, in [27, 1], Freeman et al. has shown that the performance with non-overlapping local selections is limited. Different from the NE methods, they consider the overlapping case by a two-layer MRF topology as shown in Fig. 2.3. Here, the proximity between the LR observations and the database patches are taken into account (refers to the likelihood term) in addition to the agreement between neighboring HR patches (refers to the regularization term) by

$$\hat{I}_H = \arg \max_{I_H} \left\{ \prod_i \Phi(I_H^i, I_L^i) \prod_{j,k} \Psi(I_H^j, I_H^k) \right\}, \quad (2.14)$$

where  $\Phi$  and  $\Psi$  are compatibility functions, and  $i, j, k$  are patch indexes. Hence, rather than concentrating on the true unknown image (as in NE methods), the focus



**Figure 2.3:** Graphical model for the MRF model used in [1] to define the posterior distribution of the solution space.  $\Phi$  and  $\psi$  are referred to compatibility functions and used to model local correlations.

is on the network interpretation of the data. More clearly, this interpretation refers to discovering the nearest-neighbors that survive a Bayesian Belief Propagation (BBP) algorithm and using those in the formation of the solution.

In [21], Tappen et al. has developed the MRF structure, given in (2.14), by using a computationally efficient representation. This compact representation allows a small number of discrete states to represent the local image patches. Each node in the graph denotes the index of the best regression function, chosen from a set of candidates. The best regressors at each patch are determined by using Belief Propagation. After determination of the regressor assignments for each HR patch, the missing points are synthesized by using these regressors. Note also that this new representation requires the transformation of the compatibility functions,  $\Phi$  and  $\psi$ , onto the new subspace.

A close idea to this transform domain representation is using the epitomes, which were first introduced in [25]. The epitome of an image is its condensed version containing the essence of the textural and shape properties. As in Tappen's method in subspaces [21], the size of the epitome is considerably smaller than the size of the image it represents, but the epitome still contains most constitutive elements needed to reconstruct the image. Epitomic representation provides two apparent advantages against [21]; first, it enables to use varying size image patches (this allows for model continuities better), and second, it learns from observation. Thereby, the estimation becomes more correlated with the observation. On the other hand, these advantages

come with considerable increase in computational load and lack of control. A close variant of this epitomic analysis with BBP has also been described in [56].

Baker and Kanade [10] have practiced a semi-parametric model called "reconscognition" by using an explicit regularization expression that requires proximity between the spatial derivatives of the unknown image to those of the found examples. For each pixel an example is identified by a pyramidal derivative set of features, and all these forces are merged into one global posterior distribution.

Example-based regularization is an effective technique for the single-frame SRR problem. However, there are still a number of issues needed to be considered. For instance, the proposed algorithms remain local, as they do not consider the unknown image as a whole. Moreover, the choice of patch size is not trivial. Choosing a very small patch size may cause the co-occurrence prior to be too weak to regularize the solution space sufficiently. Oppositely, too large of a patch size may lead to no adequate examples in the database. Moreover, how to choose the database is another question needed to be answered. Different images have different statistics and thereby need different databases. Also, the heavy computations required both in training and testing could be a difficulty for practical applications.

### **2.3 Heuristic Techniques**

These approaches provide solutions based on some observations, and so their theoretical foundation is a little bit less than the other approaches. However, they can produce satisfactory results for practical cases.

The Locally-Adaptive Zooming Algorithm (LAZA) has been introduced in [30] and uses a set of simple rules to extract information about discontinuities or sharp luminance variations. The algorithm is performed in four steps, and in the final step four surrounding pixels of the undefined pixel are combined to form the value of this pixel. To preserve more detail, the four pixel values are not simply averaged to form the new value; instead a histogram-like method, where the median of the bin is usually taken, is employed..

New Edge-Directed Interpolation (NEDI) [57] uses the duality between the covariances in LR and HR. The covariance between neighboring pixels in a local

window around the LR source is used to estimate the covariance between neighboring pixels in the HR target. Thus, the covariance value is used as the optimal way of blending the four diagonals into the center pixel.

In another approach (as in [29, 28, 58]), averaging pixels across boundaries is avoided by storing additional data in the form of discontinuity graphs. The effectiveness of this method can be seen best on images having a strict piece-wise planar nature, such as linear profiles separated by strong intensity jumps. In [29], the map is obtained by performing a segmentation algorithm in a pre-processing stage. A similar idea has been reported in [28] where rather than using predefined maps, a segmentation map is simultaneously learned. An improved idea is discussed in [58] by incorporating Hidden Markov Models (HMM) to extract the segmentation map. Note that this kind of goal-oriented approach is especially useful when the SRR is considered as a predecessor step of another application in a wider system (as in [59, 18]).

Smart Interpolation by Anisotropic Diffusion (SIAD) [60] uses other anisotropic diffusion algorithms as part of its three-step structure. The first step consists of enlarging the image beyond the required resolution using simple analytical interpolation. Next, an anisotropic diffusion is performed, and finally the image is reduced back to the desired size by averaging.

The main idea in Edge-Frame Continuity Modeling (EFCM) [20] is that given some local edge-related parameters (such as closeness, magnitude and scattering) extracted from the LR image for each pixel, it is possible to estimate the expected local intensity continuity observed at HR. For each pixel, these features are matched with the gradient value of the corresponding HR image at that pixel. Later, for each combination of the features a Gaussian model is built and stored in the EFCM-table. Though EFCM is good at preserving global continuity, the results suffer from the faint details and looks machinery. Moreover, finding the features for each pixel makes the solution too expensive.

These heuristic techniques may produce visually appealing images in some practical cases, but their theoretical weakness undermines their efficiency. Most of these techniques require some individual processing for each pixel, as in some interpolation methods using explicit classifiers. In real-world applications, even for mid-scale



images, this processing may cause serious difficulties. Also, as in EFCM [20], non-linear and non-convex global optimizations may be encountered (then the solution is approximated by mid-paths, such as smaller local convex optimizations).

## 2.4 Multi-Frame Super-Resolution Techniques

Having more than one observation means basically having more data about the solution, which can be utilized to restrict the solution more.

The main idea in multi-frame SRR solutions is fusing the content that is slightly different. This requires that the observations should be sub-sampled as well as shifted with sub-pixel precision. Thus, an observation cannot be obtained from the others, and each such distinct information can be exploited to construct an image in higher quality. The mathematical representation of the problem can be easily derived from the single frame case as

$$\hat{I}_H = \arg \max_{I_H} \prod_k^K p(I_L^k | I_H) p(I_H), \quad (2.15)$$

where  $K$  is the number of observations. It is possible to investigate the multi-frame SRR techniques in two main groups (frequency-domain and spatial-domain methods) depending in which domain they represent the images.

### 2.4.1 Frequency domain methods

Frequency-domain (FD) SRR methods typically rely on familiar Fourier-transform (FT) properties, especially the shifting and sampling theorems. This strict relationship to FT properties precludes the use of general observation and motion models with this type of representation. Basically, there are two approaches; one assumes noiseless environments for its observation model and derives an analytical solution [11], the other considers the ambiguity stemming from noise and uses numerical techniques by enforcing constraints to relieve this ambiguity [35, 61]. Common to both approaches is that they are based on the relation between the DFT coefficients of the observations and sampled Continuous Fourier Transform (CFT) coefficients of the HR image as

$$\mathbf{Y}_k = \Lambda \mathbf{X}, \quad (2.16)$$

where  $\mathbf{X}$  is a column vector consisting of the samples of the unknown CFT of the continuous HR image, and  $\Lambda$  is a matrix, which relates the DFT of the  $k_{th}$  LR image  $\mathbf{Y}_k$  to the samples of the continuous HR image. Therefore, the reconstruction of a desired HR image requires us to determine  $\Lambda$  and solve this inverse problem. For a noiseless case the solution is simply the multiplication of the inverse  $\mathbf{Y}_k$  and  $\Lambda$ . On the other hand, when noise or blurring is considered, the least-squares like numerical solutions are referred.

Theoretical simplicity is a major advantage of the frequency domain approach. That is, the relationship between LR images and the HR image is clearly demonstrated in the frequency domain. However, the observation model is restricted to only global translational motion and Linear-Shift-Invariant (LSI) blur. Due to the lack of data correlation in the frequency domain, it is also difficult to apply the *a priori* knowledge, given in spatial domain, for regularization.

Note also that although theoretically it is not different, as in [62], Discrete Cosine Transform (DCT) can also be used instead of DFT. Thus, the memory requirements and the computational load can be reduced.

### 2.4.2 Spatial domain methods

Different from the frequency-domain methods, the known registration assumption is relaxed in spatial domain. Due to the rich modeling capability, a wide variety of observation models can be considered. Though registration can be incorporated into the solution, generally it is quite hard to estimate the exact registration parameters since real word images have spatially varying complex geometric deformations. Ideally, these model parameters should be found for each pixel, but it is not realistic, and piecewise-homogeneity assumption is made.

Since the registration is an ill-posed inverse problem, the parameters are approximately found and this additional ambiguity makes the SRR problem harder. The most common trend in multi-frame SRR is to make registration separately [63] and utilize one of the single-frame SRR techniques by using all registered observations [16, 64]. There are also blind techniques, such as [15, 19], handling both of these problems

simultaneously. The registration parameters are combined with the HR data and are estimated in a coordinated manner.

## 2.5 Discussion

This investigation on the literature reveals that as the adaptation increases, the quality of the reconstruction gets higher. Natural images have dispersed settlement in the space, therefore the methods are expected to be adaptive enough to capture different characteristics of the imaging space. However, the increase in adaptation comes with an additional cost in computational effort. That means the relation between the quality and the complexity of the SRR methods is conflicting. For instance, the basic interpolation methods are relatively simpler than the regularization-based methods; on the other hand, the quality of their reconstruction is not as good as the regularization results. Despite this trade-off between quality and computational efficiency, it is expected that an ideal SRR method should maximize both. The ongoing research looks for such efficient techniques.

We have identified three main factors determining the efficiency of an SRR method. These are; the complexity of the reconstruction expression, the analysis power of the imposed feature set and the contribution of the data source used to extrapolate.

- **Complexity:** Computational complexity of the solution maybe the most important factor for practicality. This is because the simple kernel interpolation techniques are still the most common methods used in industry. While evaluating the complexity of a solution, several factors should be considered, such as theoretical foundation, learning (offline processing) complexity and inference (online processing) complexity. Among these factors the worst-case performance of the inference is the main determinant in most cases. When we looked at the past techniques, except for some frequency domain methods, the reconstructions mostly end up with an optimization problem. Some of these problems are convex, so tractable, while others are not. For instance, heuristic approaches and non-convex sampling-based algorithms (such as [22]), are not tractable and they are treated as if convex under some conditions. Among the convex structures, the methods having quadratic structures, such as Tikhonov Regularization, are the most

advantageous ones. It is possible to achieve the reconstruction analytically in these systems. Moreover, though they are not as convenient as the quadratic ones, the non-quadratic generalized least-squares type approaches [32, 15, 14, 65] have also mature techniques. Among these non-quadratic solutions, the unconstrained linear cost functions are especially advantageous against the constrained ones (like sparse coding).

- Features: As mentioned before, the blocking artifact problem is caused by separate vertical and horizontal treatment of the images. The limited number of features is not sufficient to analyze the whole content of an image. A wealthier set of features, consisting of intermediate orientations and multiple scales, should be incorporated to overcome this problem. Though using more features is desired, it also requires special attention on the computational load created and the artifacts caused by exceptions. A typical decision of a filter-set would include the determination of the following 4 parameters.

- Orientation: The horizontal and vertical orientation of the re-sampling kernels are unable to recognize or follow diagonal lines and this inevitably causes blocking. As a remedy, researchers provide steerable filters [66, 67] to increase the number of orientations treated.

Let us write the  $n_{th}$  derivative of a Gaussian at an angle  $\theta$  as  $G_n^\theta$ . Then, the first-order derivative in the  $x$  direction will be represented as  $\frac{\delta G}{\delta x} = G_1^0$ , and similarly in the  $y$  direction will be  $\frac{\delta G}{\delta y} = G_1^{\pi/2}$ . As shown in [66], the derivatives at intermediate orientations can be written as

$$G_1^\theta = \cos(\theta)G_1^0 + \sin(\theta)G_1^{\pi/2}. \quad (2.17)$$

Since  $G_1^0$  and  $G_1^{\pi/2}$  span the set of  $G_1^\theta$  filters, they are called basis filters for  $G_1^\theta$ . The  $\cos(\theta)$  and  $\sin(\theta)$  terms are the corresponding interpolation functions for those basis filters. Because convolution is a linear operation, we can synthesize an image, filtered at an arbitrary orientation, by taking linear combinations of the images filtered with  $R_1^0 = G_1^0 * I$  and  $R_1^{\pi/2} = G_1^{\pi/2} * I$  as

$$R_1^\theta = \cos(\theta)R_1^0 + \sin(\theta)R_1^{\pi/2}. \quad (2.18)$$

Although the illustration of the steerability is given on Gaussian filters, it is possible to generalize it to other filters, such as wedge filters as in [67]. Due

to their computational simplicity, in our experiments, we have also preferred Gaussian steerable filters when needed.

Note also that in [14] Farsiu et al. has employed another way of having steerable filters by using hand-crafted shift operators, called Bilateral Total-Variation. But, the computational load of this technique would be greater since an excessive number of online convolutions are required.

- Size: Because we develop generic solutions without having any information about the scale of the input image, we should provide a multi-scale environment where details on any scale can easily be detected. The common approach used for achieving multi-scale filters is pyramids, in which the filter size equally decreased at each level to the top [68, 10].
- Type: In SRR, our main consideration is the reconstruction of the image details, consisting of the high-frequency (HF) content, because it is assumed that the low-frequency (LF) content is incorporated with the fidelity constraint. So, while designing image priors, generally the missing HF components are considered and high-pass (HP) filters are used. Derivatives are the most popular HP filters and are designed as various order Gaussian derivatives up to the 4th order. Although the higher order derivatives provide more HF content, for noisy images they may cause artifacts. In addition to Gaussian derivatives, other edge filters and bar filters are also commonly used [27, 21, 22, 23]. Furthermore, depending on the characteristics of the problem, using specific feature detectors can also be quite helpful. For instance in [69], Torralba et al. introduces the object specific filters through the Bag-Of-Words framework.
- Number: This decision is very much related with the computational power available because each filter requires an image size convolution operation. However, there are some exceptions as in Freeman’s steerable filter design [66], where the remaining filters are obtained by interpolating basis results without doing more convolution.

One popular choice is to have up to 2nd order derivatives in at least 6 orientations (spanning  $0 - \pi/2$  interval), and on 3 different scales.

- **Reference Data Source:** Extrapolation of the right image content is quite difficult. Due to heterogeneous behavior of the natural images, the observation-based constraints are not enough to reach realistic reconstructions. So, trusted data sources are required to be able to incorporate missing image details. Based on the above literature review, it is possible to categorize the design attempts for the data source as:
  - **Learned Patch Dictionaries:** A set of codewords are learned either for a specific image domain or for the whole image space. The reconstruction is built directly by incorporating appropriate selections from this dictionary [26, 24, 21, 27, 22, 25, 10]. This idea makes a questionable assumption that the whole image space can be completely represented by a finite number of samples. Another critical problem with the idea is that the expected discontinuity comes with the incorporation of statistically independent components. Some of these studies [49, 33] assume complete independence among the dictionary elements, and others [1, 10, 21, 22] incorporate statistical dependencies among the dictionary elements during the selection process.
  - **Analytic Dictionaries:** Rather than directly using the learned dictionaries, some researchers project the image domain onto narrower subspaces and build the dictionary with less variety. Dictionaries of this type are based on some mathematical models and characterized by analytical transformations, such as wavelets, curvelets, contourlets, shearlets and bandelets [70, 71, 46]. The idea seems more efficient than working in higher-order pixel dictionaries, especially when their fast implicit implementations are considered. However, the results generally suffer from an increase in representational ambiguity since the transformations are not completely lossless.
  - **Statistical Image Models:** There are attempts to build image priors by defining density functions for the imaging space, though no regularity in natural image space has been discovered yet. However, the idea could work well especially in constrained image spaces. In [52, 22, 3, 72] researchers have introduced density functions with high representational power in various constrained image domains.

All these data source designs are far from providing the expected contribution. The huge dimensionality of the imaging space and the limited amount of resources available make the problem of building a representative data source quite hard. In this thesis, we have followed a different way to capture the missing image details and suggested using some reference or template images in HR instead of modeling. The main premise of this idea is; "gathering global continuity and realistic HF content" could only be possible by having a strong idea/experience about the content. At that point, using a structurally and semantically close reference image can represent this prior experience. Note that, since lots of mismatches are expected, only the relevant details should be considered by using intelligent techniques.

After the investigation of the past works and the observations given above, the following three motivations have been guiding our research to have efficient SRR methods:

- Having quadratic objective functions in optimization.
- Using data sources which can provide globally consistent and realistic details.
- Utilizing a wealthier set of features (at least more than horizontal and vertical derivatives) to extract different characteristics of the images.





### 3. ROBUST SUPER-RESOLUTION

As discussed in the previous chapter, modeling the natural image space is difficult because the heterogeneous nature of the images requires individual treatment of the local regions. Difficulties in representing images with complex stochastic models can be overcome by converging with the simpler deterministic structures as in anisotropic diffusion<sup>1</sup>.

In order to achieve adaptive treatment of local image regions, the simplest approach is anisotropic diffusion, where a selective treatment is employed by adjusting the weight of the imposed model on local regions. In [73] Perona and Malik provide the pioneering use of anisotropic diffusion in image processing literature by removing noise from the noisy image. The image is modified iteratively by

$$I[s]^{t+1} = I[s]^t + \frac{\lambda}{|n_s|} \sum_{p \in n_s} g(\nabla_{s,p}) \nabla_{s,p}, \quad (3.1)$$

where  $g(x)$  refers to the diffusion function (also called the evaluation or adaptation function),  $I$  is the discretely sampled image,  $n_s$  represents the spatial neighborhood of a pixel  $s(x,y)$ ,  $\nabla_{s,p}$  is the spatial derivative, and  $|n_s|$  is the number of neighbors around  $s$ . Qualitatively, the effect of anisotropic diffusion is to smooth the original image while preserving brightness discontinuities.

In addition to this pioneering interpretation (with partial differential equations) of anisotropic diffusion in image processing, later it was also interpreted from different perspectives, such as bilateral filtering [74], local mode filtering [75] and robust statistics [32]. Among these interpretations, we employ the robust statistics within this chapter.

---

<sup>1</sup>In [13], Elad et al. shows that adaptive filtering and anisotropic diffusion converges for various image processing tasks, so in the rest of this chapter they are used interchangeably.

The specific use of anisotropic diffusion in SRR is mostly in the form of generalized least squares, which is given as

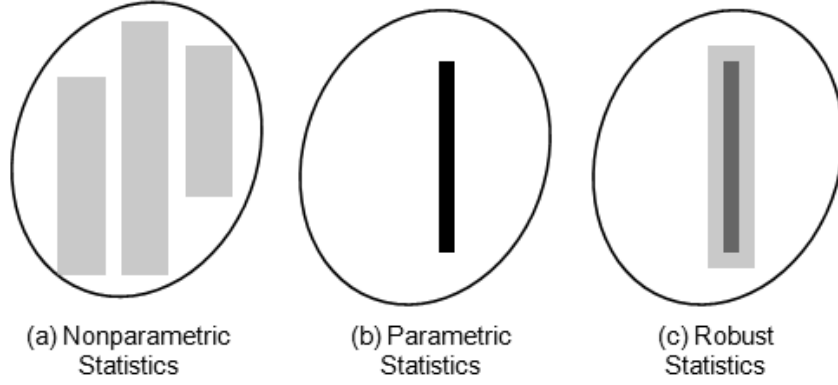
$$\hat{I}_H = \arg \min_{I_H} \{ \|I_L - HI_H\|_2^2 + \lambda \rho(\mathbf{T}I_H) \}, \quad (3.2)$$

where the regularizer is imposed after the evaluation with the  $\rho(x)$  function, which works as the  $g(x)$  function in (3.1). The choice of  $\rho(x)$  can greatly affect the extent to which discontinuities are preserved. In [32], Black et al. provides a statistical interpretation of anisotropic diffusion, specifically from the point of view of robust statistics. Robust error norms can be used as  $g(x)$  to maximize the preservation of the edge regions while imposing a smooth image. These functions are able to minimize the effect of the gross outliers. In imaging, the outlier does not only mean the additive observation noise, instead it is used for all sorts of discordant observations caused by: observation noise (like the mixing of two signals), measurement errors (such as quantization) or limitations in data models.

In this chapter, we have investigated the application of robust error norms for the SRR problem in the form of anisotropic diffusion. We have proposed an efficient SRR solution for the cases where training is not possible due to either lacking enough data or making the solution generic. Specifically we have used a special robust function, called the Welsch norm, to adjust the diffusion rate. The Welsch norm has better edge-stopping utility than other robust estimators. Also, its partially-quadratic structure guarantees the unique solution with gradient descent methods. In addition to these, we have employed a data source to better regularize the solution. By these additional constraints we could clone globally consistent image details, which cannot be retrieved with model based or observation dependent constraints. To show the effectiveness of the proposed reconstruction scheme, we have provided a set of experiments with different features and reference images.

### 3.1 Robust Statistics

In a broad informal sense, Hampel et al. [2] defines Robust Statistics as a collection of related theories, concerning with the fact that many assumptions commonly made in statistics (such as normality, linearity, independence) are at most approximations to reality. In addition to outliers, another important reason for that is the deviations



**Figure 3.1:** The space of all probability distribution on a sample space (denoted with the ellipsoid). (a) Non-parametric statistics: allow almost all possible distributions (restriction is quite limited and this ignorance is represented with an interval) (b) Parametric statistics: define strictly determined distributions (represented with a straight line). (c) Robust Statistics: define a neighborhood of strict parametric statistics by allowing slight fuzziness [2].

between the empirical character of the models and the approximate character of the theoretical models (e.g. non-uniform natural image space is approximated with a uniform image model in Tikhonov regularization). Given this situation, the problem with the theories of classical parametric statistics is that they derive the optimal procedures under the exact parametric models, but say nothing about their behavior when the models are only approximately valid. Even, the nonparametric statistics do not specifically address this situation.

At that point, robust statistics allow a full neighborhood of a parametric model; thus, being more realistic and yet, apart from some slight fuzziness, providing the same advantages as a strict parametric model (see Fig. 3.1).

In literature several approaches to robust estimation have been proposed, including M-estimators, R-estimators <sup>2</sup> and L-estimators <sup>3</sup> [76]. However, M-estimators now appear to dominate the field as a result of their generality, high breakdown point, and their computational efficiency. M-estimators are a generalization of Maximum Likelihood Estimators (MLE) where we try to maximize the total probability  $\prod_{i=1}^n f(x_i)$  over all data points or equivalently minimize  $\sum_{i=1}^n -\log f(x_i)$  [9]. In [76], Huber has proposed to generalize this to the minimization of  $\sum_{i=1}^n \rho(x_i)$ , where  $\rho(x)$

<sup>2</sup>An r-estimator is an estimator based on rank test.

<sup>3</sup>An L-estimator is an estimator which equals a linear combination of order statistics of the measurements.

is some function, not necessarily a distribution as in  $f(x)$ . Minimizing  $\sum_{i=1}^n \rho(x_i)$  can often be done by differentiating  $\rho(x)$  and solving  $\sum_{i=1}^n \psi(x_i) = 0$ , where  $\psi(x) = \frac{\delta \rho(x)}{\delta x}$  is called the *influence function*.

Huber provides a list of standard properties that a reasonable objective function  $\rho(x)$  of an M-estimator must satisfy:

- $\rho(x) \geq 0$ ,
- $\rho(0) = 0$ ,
- $\rho(x) = \rho(-x)$ ,
- $\rho(x) \geq \rho(y)$  for  $|x| \geq |y|$ ,
- $\rho(x)$  is differentiable.

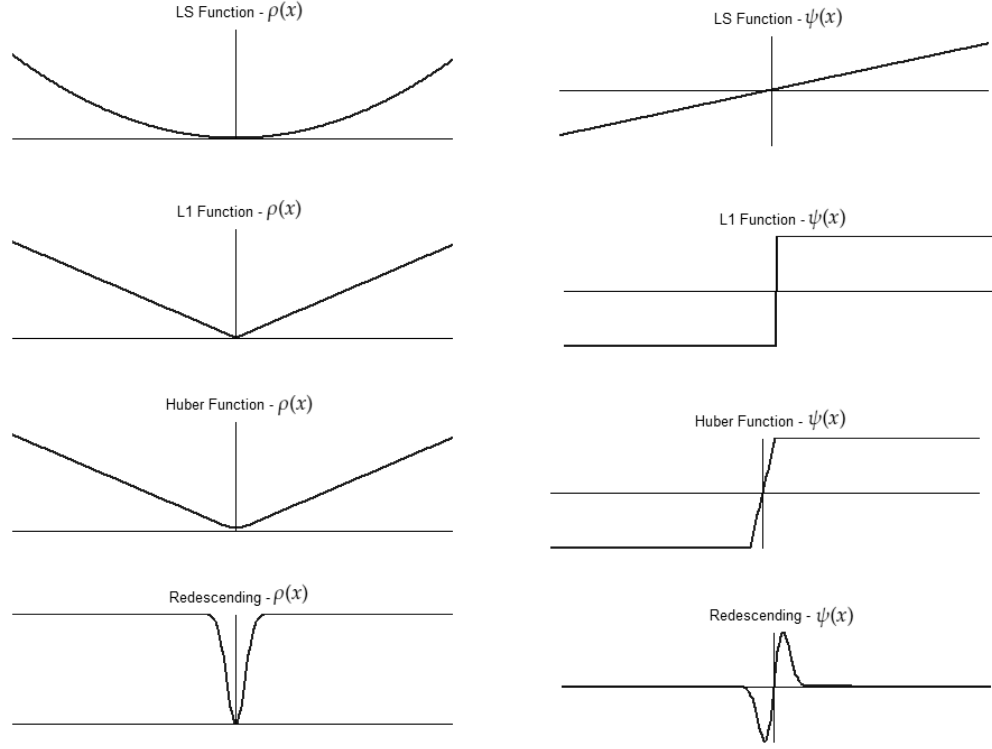
In Table 3.1 some popular  $\rho(x)$  functions, satisfying these conditions, are listed.

**Table 3.1:** Some popular M-Estimators.

Estimator	$\rho(x)$	$\psi(x)$
$L_2$	$x^2/2$	$x$
$L_1$	$ x $	$sign(x)$
Huber's MinMax [76]	$x^2/2$ for $ x  \leq c$ $c( x  - \frac{c}{2})$ for $ x  > c$	$x$ for $ x  \leq c$ $c(sign(x))$ for $ x  > c$
Redescending Estimators (as an example Welsch Norm)	$\frac{c^2}{2}[1 - e^{-(x/c)^2}]$	$xe^{-(x/c)^2}$

The  $\rho(x)$  and  $\psi(x)$  functions for these estimators can be shown in Fig. 3.2. The non-robust least-squares (LS) estimate (namely the  $L_2$ -norm) is very sensitive to outliers, because the influence function increases linearly and without bound. That means, when the values have different characteristics (that is coming from different populations, e.g. pixels across a boundary) the mean is not representative of either population, and the image is blurred. However the other norms in Fig. 3.2 are robust and limit the effect of the outliers on the solution. When the value of a sample is beyond a limit, the influence of that sample is fixed, even reduced.

Although the  $L_1$  norm is robust, it is criticized for producing estimates with a higher variance than quadratic norm functions. It is worth noting that  $L_p$  estimators ( $1 \leq$



**Figure 3.2:**  $\rho(x)$  and  $\psi(x)$  functions of some popular M-estimators.

$p \leq 2$ ) do not require a scale,  $c$ , estimate and hence have an advantage of one degree of freedom (possibly will be useful when the available resources are limited for learning).

The common feature of the remaining two functions of Table 3.1 is their quadratic treatment of error up to a threshold and then getting into a saturation stage to treat the remaining values almost uniformly. This kind of behavior perfectly matches the nature of the imaging space. Among these robust estimators, within this section we are going to be specifically interested in re-descending M-Estimators since they completely reject the outliers exceeding a certain limit.

### 3.1.1 Re-descending M-estimators

Re-descending M-estimators are those M-estimators that are able to reject extreme outliers completely. In addition to the standard properties of M-Estimators, a re-descending M-estimator should also satisfy the following condition

- $\lim_{r \rightarrow \infty} \psi(r) = 0$ , where  $\psi(r) = \frac{\delta \rho}{\delta r}$ .

Several choices of  $\rho(x)$  functions, having a re-descending  $\psi(x)$  function, have been proposed in literature. Some popular re-descending M-estimators are listed in Table 3.2, and graphically shown in Fig. 3.3.

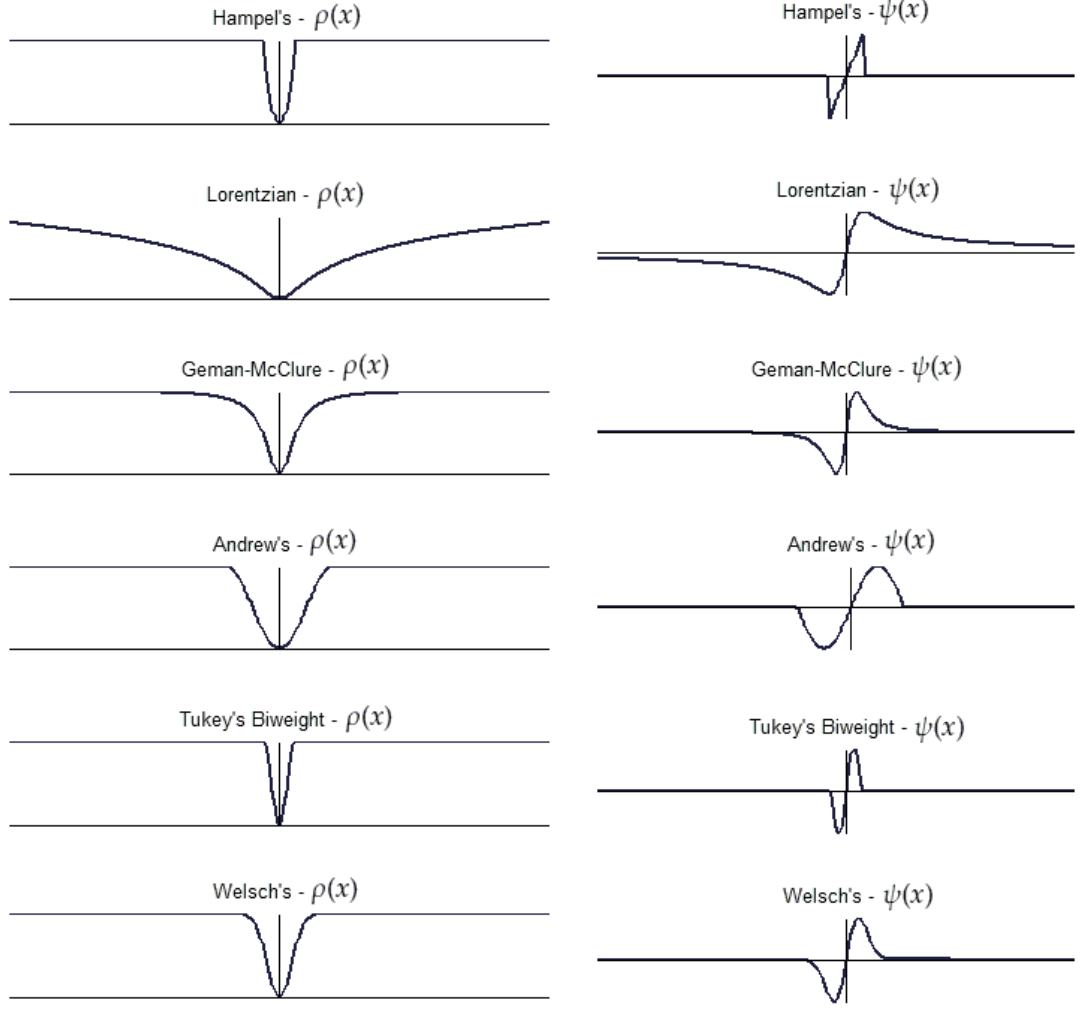
**Table 3.2:** Some popular Re-descending M-Estimators.

Estimator	$\rho(x)$	$\psi(x)$
Hampel's Norm [2]	$\begin{aligned} & x^2/2 &  x  \leq a \\ & a( x  - a/2) & a <  x  \leq b \\ & a(b - a/2) + a( x  - b)(1 - \frac{2( x -b)}{c-b}) & b <  x  \leq c \\ & a(b - a + c)/2 &  x  > c \end{aligned}$	$\begin{aligned} & x \\ & a \operatorname{sign}(x) \\ & a \operatorname{sign}(x) (\frac{c- x }{c-b}) \\ & 0 \end{aligned}$
Andrew's Norm [43]	$\begin{aligned} & c(1 - \cos(x/c)) &  x  \leq c\pi \\ & 2c &  x  > c\pi \end{aligned}$	$\begin{aligned} & c \sin(x/c) \\ & 0 \end{aligned}$
Geman-McClure Norm [77]	$\frac{x^2}{2(c^2 + x^2)}$	$\frac{xc^2}{(c^2 + x^2)^2}$
Lorentzian Norm [76]	$\frac{c^2}{2} \log(1 + (\frac{x}{c})^2)$	$\frac{x}{1 + (x/c)^2}$
Tukey's Biweight [78]	$\begin{aligned} & \frac{c^2}{6} (1 - (1 - (\frac{x}{c})^2)^3) &  x  \leq c \\ & \frac{c^2}{6} &  x  > c \end{aligned}$	$\begin{aligned} & x(1 - (\frac{x}{c})^2)^2 \\ & 0 \end{aligned}$
Welsch's Norm [79, 80]	$\frac{c^2}{2} [1 - e^{-(x/c)^2}]$	$xe^{-(x/c)^2}$

Though all these re-descending M-estimators work well in detecting outliers and eliminating their influence on the estimates, their implementations are not always easy. For instance, Hampel's three part function requires the user to choose three tuning parameters, which is undesirable. Moreover, the lack of differentiability of its  $\psi(x)$  function is not ideal. The Lorentzian error norm and the Geman-McClure norm are criticized for their slow rejection rate, and because of that the outliers continue to affect the solution more than the others.

On the other hand, the remaining re-descending estimators; Andrew's sine function, Tukey's biweight and Welsch norm; are relatively more convenient to work with. They have faster decreasing rate, and immediately after a certain threshold their influences reach negligible proportions. Thus, very extreme observations are removed from the estimate. In terms of image processing, this feature is quite useful since it preserves edges while imposing generic image models. Among these three useful estimators, the Welsch<sup>4</sup> norm is much more convenient, since we can write the Welsch norm in closed

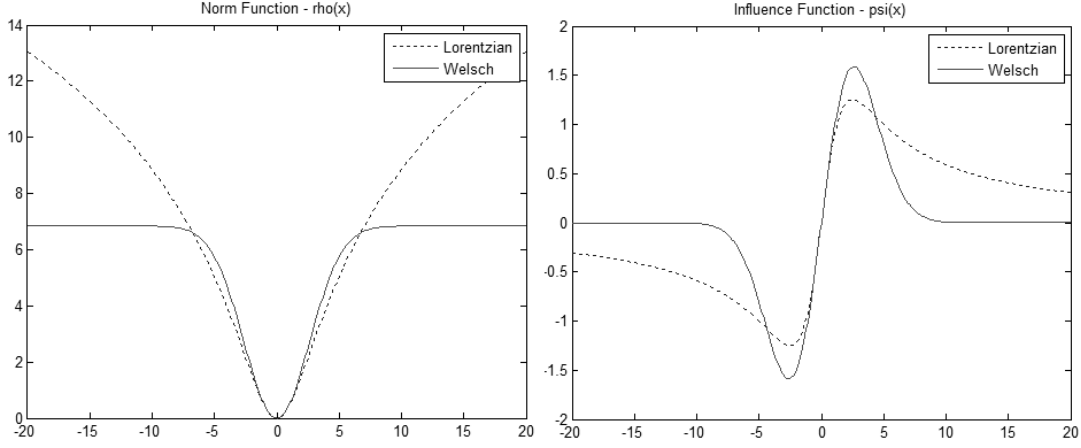
<sup>4</sup>Though it is mostly known as Welsch norm [79], a more generic class of estimators, including the Welsch norm, has been introduced as Ea-type estimators by Ramsay [80]. Note also that this robust function has been interpreted with functions having similar forms. For instance in [81] Leclerc employs  $1 - N(0, \sigma^2)$ , which has almost the same structure with the Welsch norm.



**Figure 3.3:**  $\rho(x)$  and  $\psi(x)$  functions of some popular re-descending M-estimators.

form without using an indicator function. This saves several steps in programming and provides conveniences in theoretical derivations.

In SRR literature, the re-descending M-estimators have been investigated for only the Lorentzian norm, as in [31, 65]. But the other more efficient error norms, especially Tukey's biweight and Welsch norm, have not yet been investigated. Only the Tukey's biweight function has been proposed for the denoising problem by Black et al. [32]. In that work Tukey's biweight function has been compared with the Lorentzian error norm, and it has been reported that the Tukey's biweight eliminates the noise more accurately. Despite this observation, the popularity of the Lorentzian error norm is mainly explained with its computationally efficient structure. Since it has a differentiable closed form, theoretical derivations and mathematical programming become easier. At that point, the Welsch norm not only includes these advantages but



**Figure 3.4:** Comparison of the Lorentzian and Welsch norms. Left belongs to the comparison of  $\rho(x)$  functions and right is for the comparison of  $\psi(x)$  functions.

also provides a much better adaptive treatment than the Lorentzian function. In Fig. 3.4, we compare the  $\rho(x)$  and  $\psi(x)$  functions of Lorentzian and Welsch norms. A direct comparison requires that we dilate and scale the functions to make them as similar as possible. We dilate the free scale parameter,  $c$ , of the norm functions so that they begin rejecting outliers at almost the same value.

It is clearly seen from Fig. 3.4 that the choice of the evaluation function  $\rho(x)$  highly affects the stopping behavior of the diffusion. Given a piecewise constant image where all discontinuities are above a threshold, the Welsch norm will leave the image unchanged, whereas the Lorentzian function will not.

### 3.2 SRR With Welsch Type Robust Error Norms

Traditionally, the SRR methods are interpreted as either a multiple constraint optimization, or a statistical regularization problem. However, as in [82, 13], there are works revealing the relation between these two interpretations. When appropriate functions have been used, it is possible to find the right transformations between these two interpretations. Based on this observation, we do not show special attention to this difference in view of the problem, and explain our solution in the easiest way, that is the cost function perspective. When needed, the corresponding relations can be derived by using the expressions provided in [82, 13].



The basic Tikhonov method is based on the addition of a quadratic penalty to the standard data fidelity criterion (which is also quadratic)

$$\hat{I}_H = \arg \min_{I_H} \{ \|I_L - HI_H\|_2^2 + \lambda \|\mathbf{\Gamma} I_H\|_2^2 \}, \quad (3.3)$$

where  $\mathbf{\Gamma}$  is some image feature operator; e.g. functioning as some derivative filter. The use of such quadratic, L2-based criteria for the data and the regularizer leads to the linear solution. While such linear processing is desirable, it is also limiting. In particular, when used for suppressing the effect of high-frequency noise, such linear filters also reduce the HF energy in the true image; hence, blur the details in reconstruction. Oppositely, far more powerful results are possible if non-linear methods are allowed. To allow using various types of functions in (3.3), it can be generalized as

$$\hat{I}_H = \arg \min_{I_H} \{ J_1(r_1(I_H)) + \lambda J_2(r_2(I_H)) \}, \quad (3.4)$$

where  $J_i(x)$  represent the evaluation functions for the corresponding response functions  $r_i(x)$ . In literature various combinations of  $J_1(x)$  and  $J_2(x)$  have been investigated. For instance in [14], Farsiu et al. has experimented with the combination of L2 and L1 norms as  $J_1(x)$  and  $J_2(x)$ , respectively. Similarly in [31, 65], the use of the Lorentzian norm has been investigated as  $J_2(x)$  by experimenting together with L2 as  $J_1(x)$ . In [15], the Huber function has been used as  $J_2(x)$ . Though not directly related to the SRR problem, Black et al. [32] has employed L2 norm as  $J_1(x)$  and the Tukey's biweight function as  $J_2(x)$  in the denoising problem.

Based on the discussion of the previous section, we have proposed using Welsch's re-descending norm when needed. In the following subsections we build our proposed reconstruction scheme by investigating alternative designs for the solution components (that is the response and diffusion functions).

### 3.2.1 Response functions

Both machine and human perception are performed based on the High-Frequency (HF) content of the images. This discriminative content may be disturbed by filtering, decimation and noise, as in our assumed forward model (1.7). So, heavy interest in SRR is devoted to adding these missing HF components.

However, the existing Low-Frequency (LF) content should also be preserved. It is assumed that the LR observation represents the LF content of the image to be estimated. So, the whole content of the observation should be incorporated into the solution while designing the response function  $r_1(x)$  for data fidelity. To realize this intent, we prefer working on pixel domain by measuring the distance between the LR observation  $I_L$  and the HR intermediate estimate  $I_H$  as

$$r_1(I_H|I_L) = I_L - HI_H. \quad (3.5)$$

In literature various alternative functions have been proposed such as sparse coding [26, 47] and dictionary-based representations [71, 1, 49]. However, all these attempts are based on lossy transformations in generic image spaces, and they are particularly more useful in constrained image domains.

As to the regularization term; the most widely used *a priori* information about the natural image space is the smoothness assumption [23]. Smoothness constraint provides the elimination of the unreliable HF components, and these components can be extracted via first order derivative filters in vertical and horizontal orientations,  $\Gamma_i$ . By using these features, the smoothness constraints can be designed in the form of

$$r_2(I_H(x, y)) = \sum_i^{N_s} (\Gamma_i * I_H)(x, y), \quad (3.6)$$

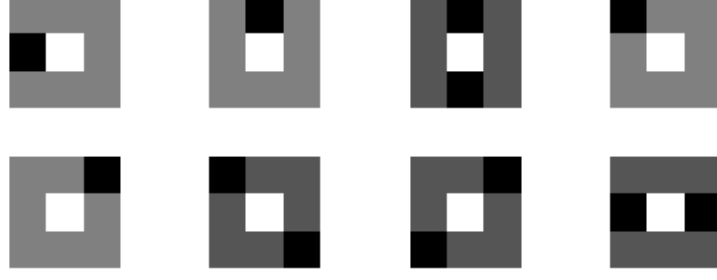
where  $N_s = 2$ , and  $\Gamma_1$  and  $\Gamma_2$  are the horizontal and vertical derivative kernels, respectively. As mentioned in Chapter 2, such kind of bilinear treatment cannot extract the content along diagonal edges and lines. In order to overcome this problem, we suggest using a wider set of filters consisting of 1st and 2nd order derivatives at 4 orientations, as shown in Fig. 3.5.

### 3.2.2 Quadratic norms vs robust norms

Though not enough, the measurement data are the main source of information. So, the solution should primarily conform with the observation. To strictly enforce these data fidelity constraints, we have employed the quadratic L2 norm as

$$J_1(r_1(I_H|I_L)) = \|I_L - HI_H\|_2^2. \quad (3.7)$$

Hence, we maximize the number of constraints for the reconstruction. In fact, this quadratic selection is very much dependent on our assumed noise model, used in



**Figure 3.5:** Derivative features used to impose smoothness in the solution. The feature set consists of 8 filters representing the first 2 order derivatives and intermediate orientations to involve also the diagonal components.

the image formation (1.7). We assume that the observation noise is additive white Gaussian, and so L2 norm provides us a safe means to do this.

As aimed for, making all the cost terms in (3.4) convex and quadratic structures simplify the optimization greatly. However, images tend to have smooth regions interrupted by sharp discontinuities and with quadratic error norms, the influence of outliers may suddenly be dominant. To avoid over-smoothing, robust error norms would be good choices (because of their natural edge-stopping functionality and computational simplicity) to estimate the piecewise-homogeneity

$$J_2(r_2(I_H)) = \sum_{(x,y) \in I_H} \sum_i^{N_s} \frac{c^2}{2} \left( 1 - \exp \left( - \left( \frac{(\Gamma_i * I_H)(x,y)}{c} \right)^2 \right) \right). \quad (3.8)$$

Generally, the robust functions do not admit closed form solutions, and often result in an objective function that is non-convex, see Figures 3.2 and 3.3. Though stochastic minimization techniques such as simulated annealing can be used with these non-convex structures, they would not be efficient enough for practical use. However if we choose a suitable robust  $\rho(x)$  function that is twice differentiable, then a local minimum can be found using deterministic continuation methods (such as the descent methods). Robust functions have scale parameters which allow the shape of the functions to be changed. By adjusting the scale  $c$  (either automatically by using the tools from robust statistics [83] as,  $c = 1.4826(MAD(\nabla I))$ <sup>5</sup>, or empirically to fit the test data more), we can make the problem convex within the accepted solution space. At that point the Welsch norm satisfies all these conditions (having closed

<sup>5</sup>MAD denotes the Median Absolute Deviation.

form, continuously differentiable, and single scale parameter) and provides superior adaptation compared to other re-descending M-estimators as shown in Fig. 3.4.

### 3.2.3 Data synthesis

When we decimate the data, as in SRR, we completely lose some portion of it. This missing data should be re-generated during reconstruction. However, as shown in the conditioning analysis of the problem [10, 12], as the magnification increases, the required number of conditions for the unique solution increases quadratically as well. So, in addition to the smoothness constraint, a trusted data source, from which reliable image details can be incorporated, should be employed to more restrict the solution space.

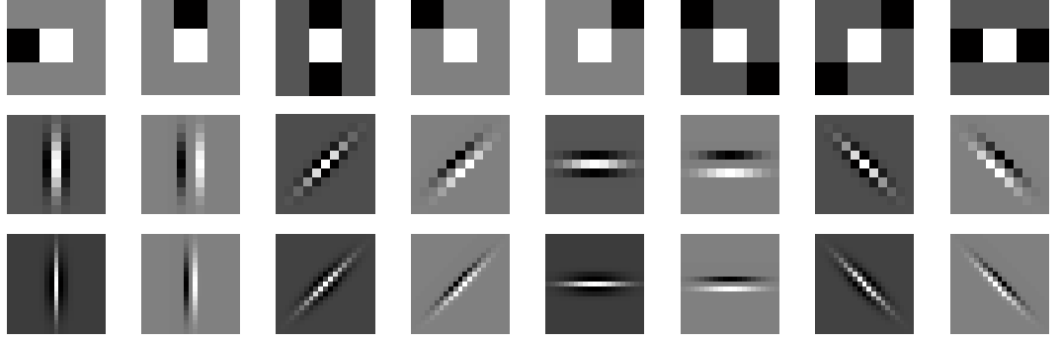
Some popular data source design attempts were reviewed in Chapter 2. The common feature of all these source definitions is to model local relations. However, due to the weak representational power the results mostly suffer from discontinuity and blurring artifacts. To remedy over-generality and discontinuity, we have followed a memory-based technique. Assuming that existing band-pass components of the observation are enough for discrimination and semantically close images have similar HF components, in a generic image database we have searched for a close match. The best matching image, called as reference, has roughly warped to the observation to alleviate the effects of geometric differences. This found reference image  $S$  has been employed as the data source to extrapolate.

During cloning we should consider these two conflicting goals simultaneously: allow copying missing HF components in an extent, and avoid copying irrelevant details. To do this, we have followed a similar approach as in smoothing constraint, and under the same robust form we have measured the distance between the HF components extracted by using oriented pyramids and multi-order derivative filters as

$$r_3(I_H(x,y)|S) = \sum_j^{N_d} ((\Gamma_j * I_H)(x,y) - (\Gamma_j * S)(x,y)), \quad (3.9)$$

$$J_3(r_3(I_H|S)) = \sum_{(x,y) \in I_H} \sum_j^{N_d} \frac{c_d^2}{2} \left( 1 - \exp \left( - \left( \frac{(\Gamma_j * I_H)(x,y) - (\Gamma_j * S)(x,y)}{c_d} \right)^2 \right) \right). \quad (3.10)$$

Thus, faint borders have been considered as outliers and they have been eliminated during reconstruction. The selected features for the cloning are shown in Fig. 3.6.



**Figure 3.6:** Edge and bar features used to extract the HF content while cloning the image details.

By the addition of the data cloning, the total solution would be a reasonable combination of these partial solutions. A total cost/energy function has been defined over the space of all possible high-resolution images by substituting (3.7), (3.8) and (3.10) into (3.4),

$$\begin{aligned} \mathbb{E}(I_H) = & \alpha \|I_L - HI_H\|_2^2 + \beta \sum_{(x,y) \in I_H} \sum_i^{N_s} \frac{c_s^2}{2} \left(1 - \exp\left(-((\Gamma_i * I_H)(x,y)/c_s)^2\right)\right) + \\ & \gamma \sum_{(x,y) \in I_H} \sum_j^{N_d} \frac{c_d^2}{2} \left(1 - \exp\left(-(((\Gamma_j * I_H)(x,y) - (\Gamma_j * S)(x,y))/c_d)^2\right)\right), \end{aligned} \quad (3.11)$$

where  $\alpha, \beta, \gamma$  are used to adjust the contribution of each cost term and satisfy  $\alpha + \beta + \gamma = 1$ . Since we have assumed that all local regions share the same set of parameters, the weighing parameters and the scale parameters of the norm functions ( $c_s$  and  $c_d$ ) have been designed empirically to generate the desired results. For the minimization of this cost function we prefer using the gradient descent optimization via the following gradient expression,

$$\begin{aligned} \nabla \mathbb{E}(I_H) = & -2\alpha H^T (I_L - HI_H) + \beta \sum_i^{N_s} \Gamma_i^T \left( (\Gamma_i I_H) \div \left( \exp\left(\frac{(\Gamma_i I_H) \odot (\Gamma_i I_H)}{c_s^2}\right) \right) \right) + \\ & \gamma \sum_j^{N_d} \Gamma_j^T \left( (\Gamma_j I_H - \Gamma_j S) \div \left( \exp\left(\frac{(\Gamma_j I_H - \Gamma_j S) \odot (\Gamma_j I_H - \Gamma_j S)}{c_d^2}\right) \right) \right), \end{aligned} \quad (3.12)$$

where  $\odot$  represents element-by-element matrix multiplication, and  $\div$  is the element-by-element division operator. Also,  $\Gamma_i$  is the convolution operator corresponding to the feature  $\Gamma_i$ , and basically it refers to the convolution of the image  $I$  with the feature  $\Gamma_i$  at all pixels;  $\Gamma_i I \sim \{\forall (x,y) \in I, \quad (\Gamma_i * I)(x,y)\}$ .

### 3.3 Experiments

As mentioned before, some re-descending M-estimators have already been proposed for the SRR problem in literature. Almost all of these works employ the same type of estimator, that is the Lorentzian norm, as in [31, 65]. Thus, we have investigated the performance of our proposed reconstruction scheme with the Welsch norm, through a comparison with the results of the popular Lorentzian norm. For the comparisons we have considered the simpler form of the reconstruction scheme as

$$\mathbb{E}(I_H) = \alpha \|I_L - HI_H\|_2^2 + \beta \sum_{(x,y) \in I_H} \sum_i^{N_s} \frac{c_s^2}{2} \left( 1 - \exp \left( - \left( \frac{(\Gamma_i * I_H)(x,y)}{c_s} \right)^2 \right) \right) \quad (3.13)$$

where the data cloning constraints are neglected. Moreover, to make the evaluations fair enough, we have scaled each function as it gets into saturation at the same rate. We have used 4 different images as shown in Figures 3.7-3.10 and the corresponding observations have been obtained by: 2x2 decimation, PSF blurring with a 5x5 Gaussian filter having the parameters  $N(0, 1)$ , and corrupting with additive white Gaussian noise having the variance  $\sigma_n = 10$ . Also the mixing weights,  $\alpha$  and  $\beta$ , have been adjusted to 0.5 for equal contribution.

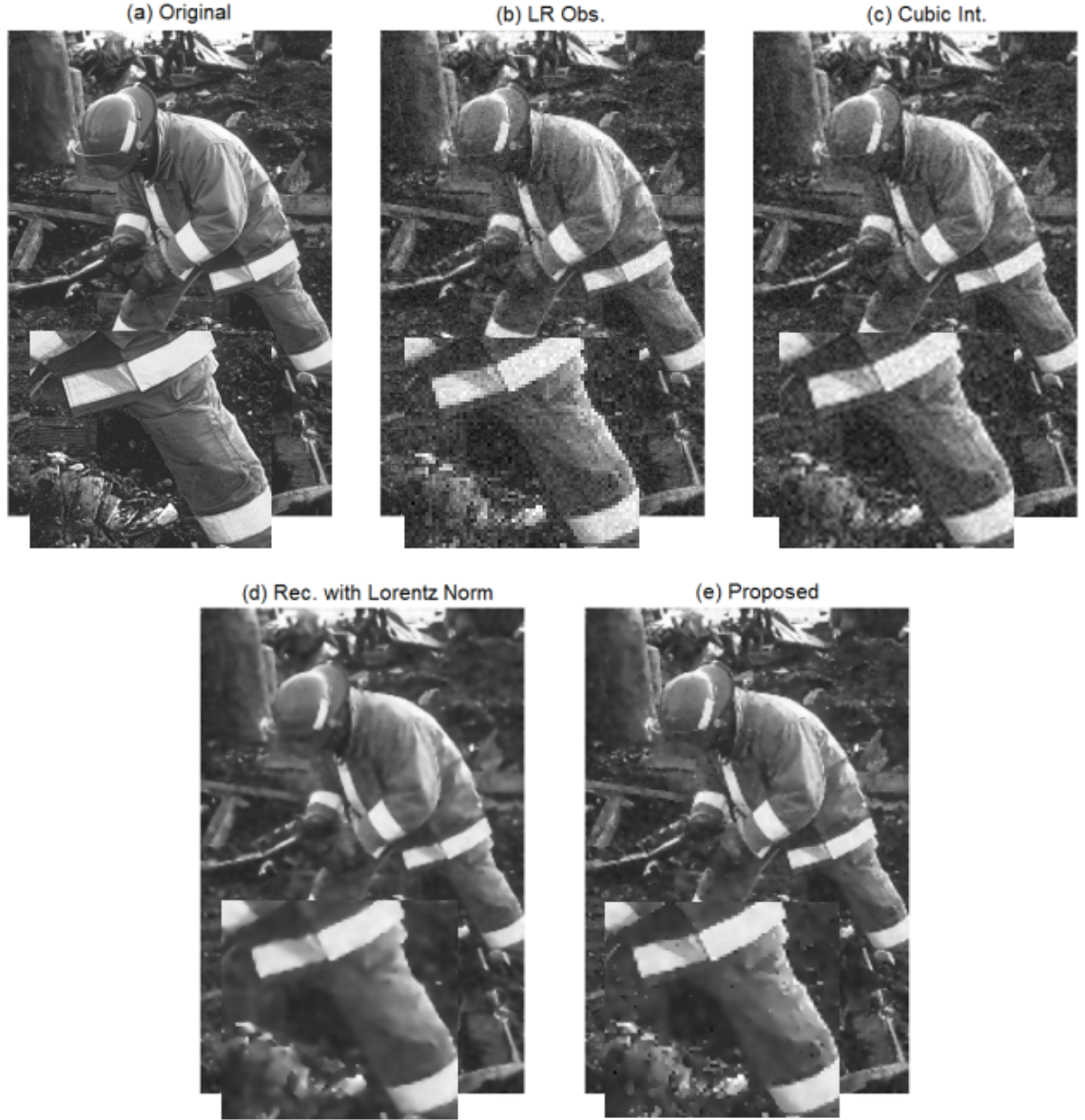
As seen from the Figures 3.7-3.10, the Welsch norm produces perceptually and quantitatively <sup>6</sup> superior reconstructions (only in Fig. 3.10 the interpolation result is competitive and this is mainly caused by the less amount of detail in the image). The results obviously show that the Lorentzian error norm has a slower transition from rejecting outliers; thus, the results get much smoother.

In order to observe the behavior of the proposed reconstruction scheme at different scales, we have performed another experiment by changing the parameter  $c$  of the Welsch function. In this experiment we have used again the simplified reconstruction expression, given in (3.13).

The results in Fig. 3.11 show that as the scale increases the saturation level of the model increases, and the behavior of the evaluation function converges to the behavior of the L2 norm. Smoothness assumption is employed almost homogeneously and the

---

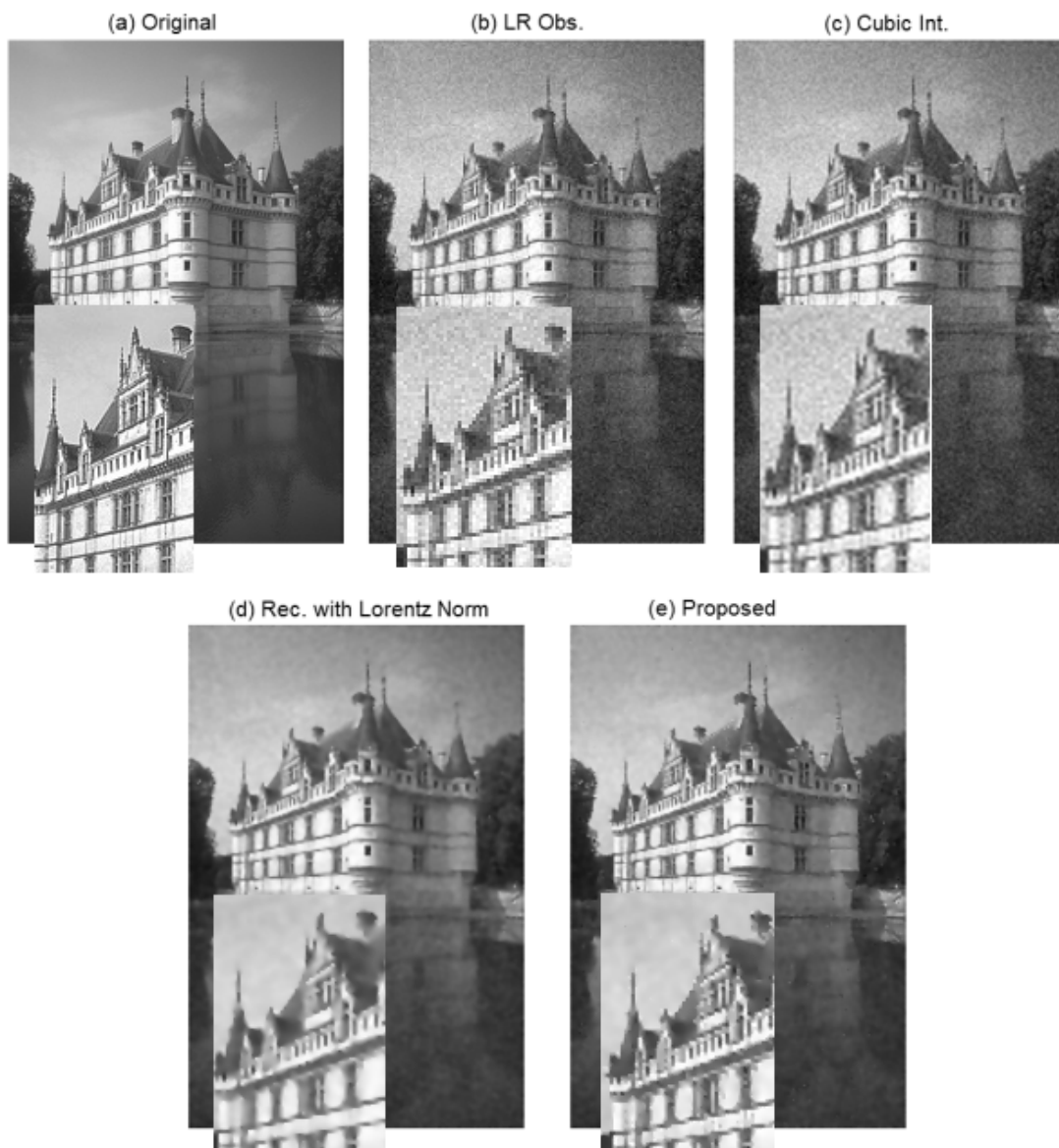
<sup>6</sup>For quantitative comparisons Root Mean Squared Error (RMSE) were used. RMSE is found as:  
 $RMS E(X1, X2) = \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}}$ .



**Figure 3.7:** Performance comparison of the Lorentzian and Welsch type M-estimators in the SRR scheme given in 3.13. (a) Original HR image (Fireman). (b) LR observation (RMSE=20.58). (c) Bicubic interpolation (RMSE=17.79). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=16.54). (e) Reconstruction with the proposed Welsch type error norm (RMSE=16.13).

image details are lost. As a result, excessively blurred images are obtained. Oppositely, at small scales, the saturation starts quite early and rewards the unwanted noisy pixels. These undesired behaviors at two extremes reveal that the scale parameter should be selected so that the evaluation function has a balanced behavior.

Up to now we have neglected the data cloning term of the proposed reconstruction (3.11). When only the smoothness constraints have been imposed, the results suffer

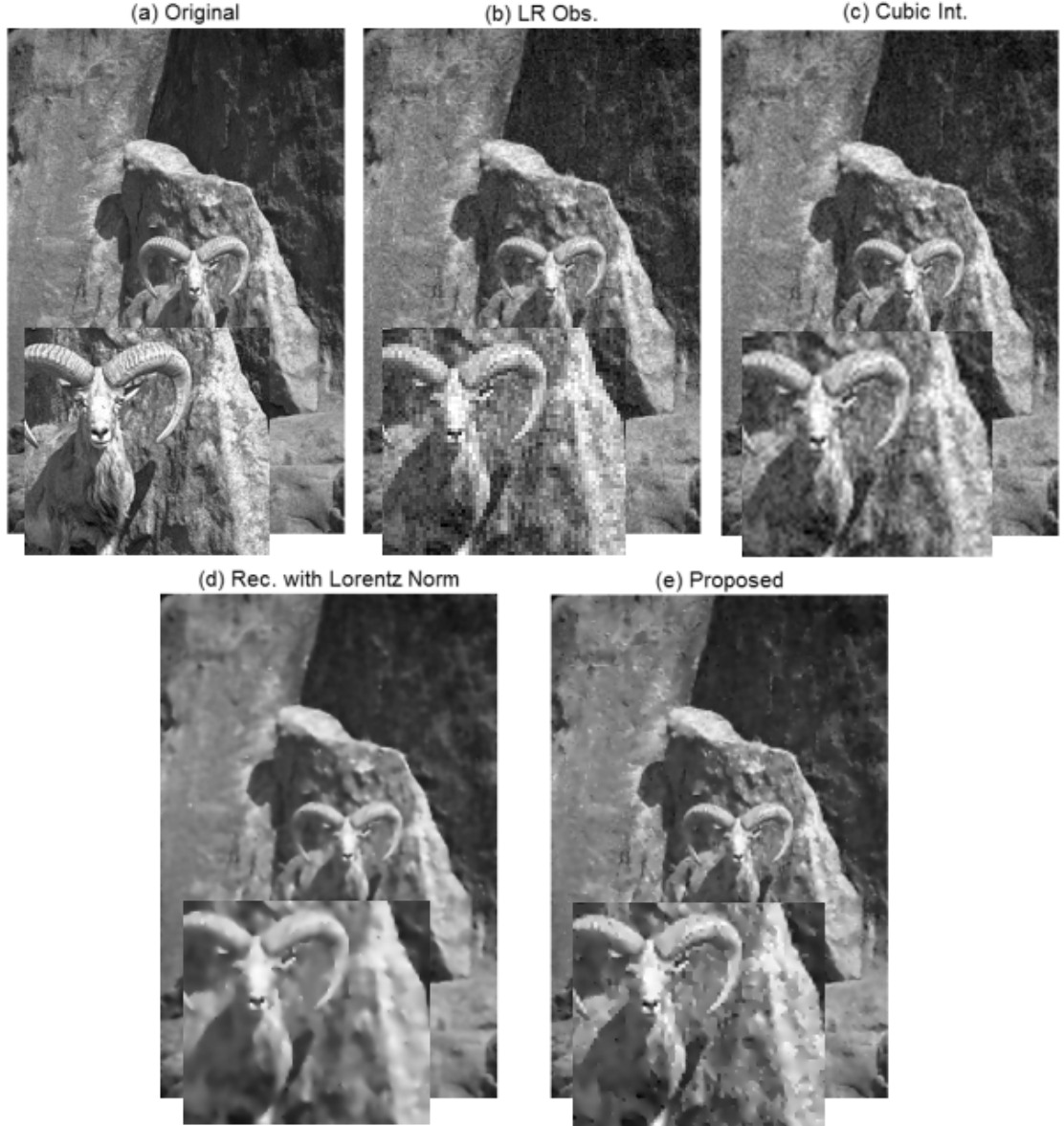


**Figure 3.8:** Performance comparison of the Lorentzian and Welsch type M-estimators in the SRR scheme given in 3.13. (a) Original HR image (Castle). (b) LR observation (RMSE=18.65). (c) Bicubic interpolation (RMSE=16.56). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=14.68). (e) Reconstruction with the proposed Welsch type error norm (RMSE=14.55).

from image details as in Figures 3.7-3.11. To overcome this problem relevant details are cloned from an outside data source, which is structurally close to the observation. In this experiment we show the contribution of this additional data cloning term.

The reference image should be aligned with the observation to maximize the contribution of these additional constraints. However, in some real world applications it could be hard to determine the exact alignment parameters, even if the images have a similar texture. Based on this fact, we present two types of experiments considering

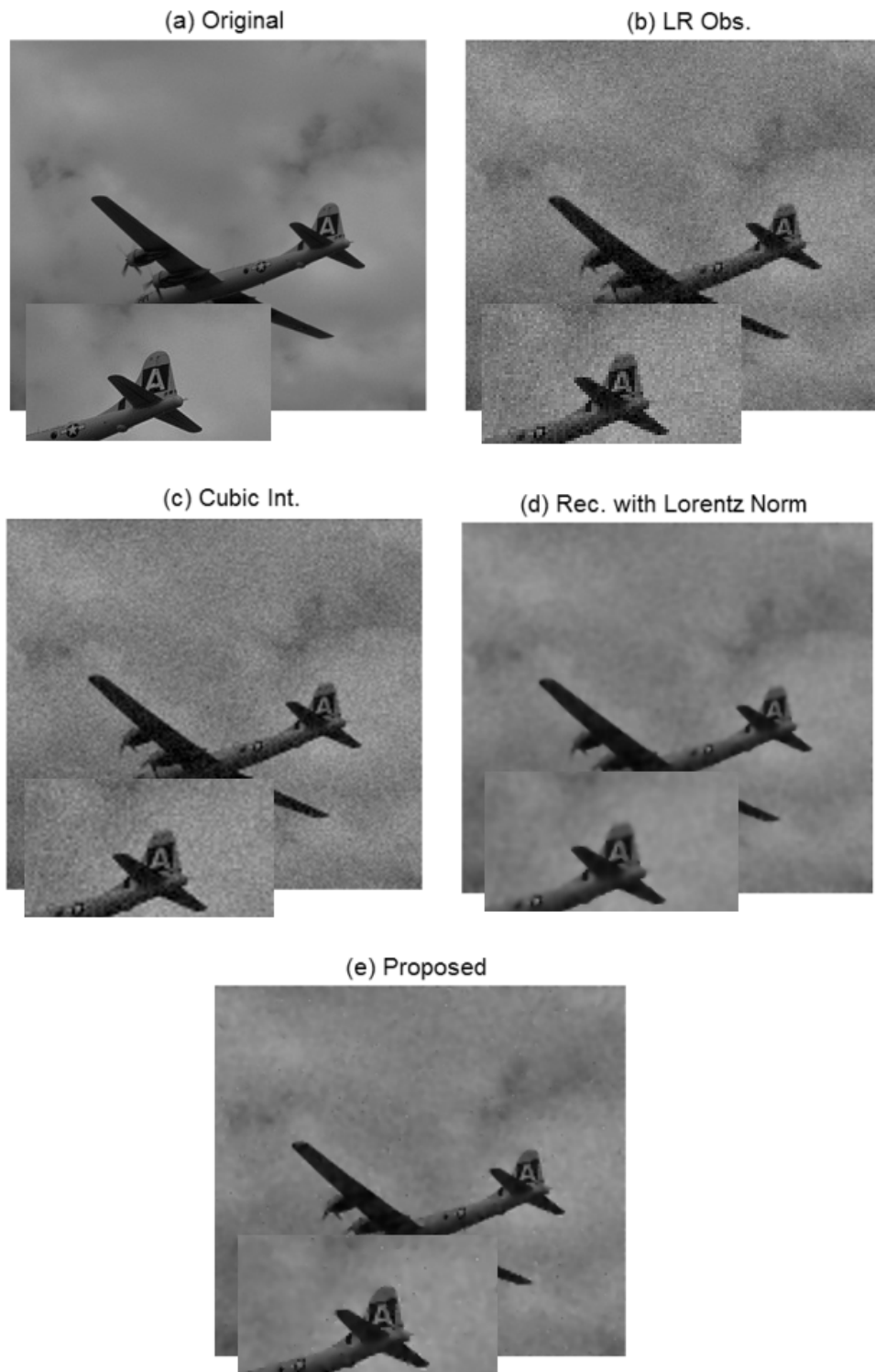




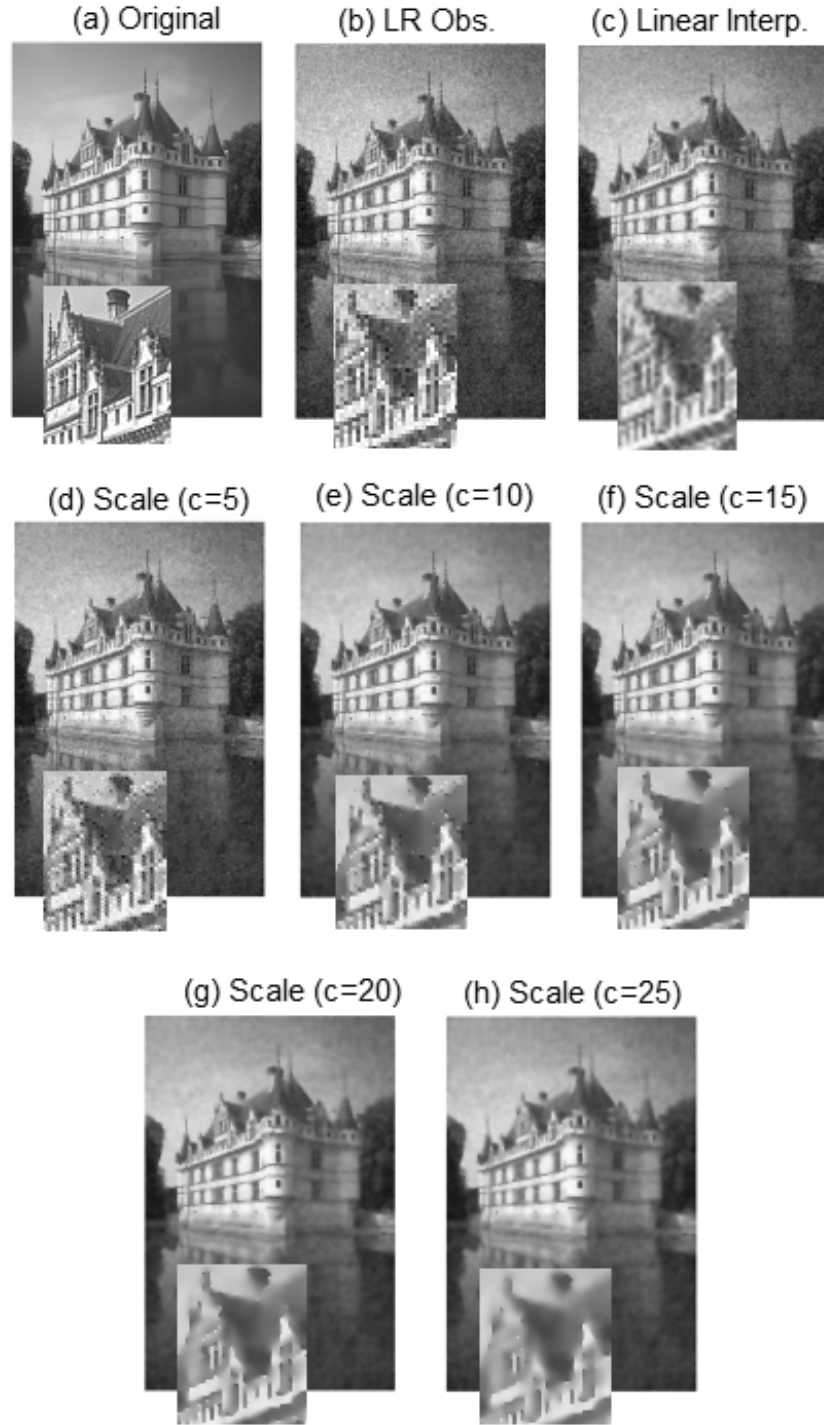
**Figure 3.9:** Performance comparison of the Lorentzian and Welsch type M-estimators in the SRR scheme given in 3.13. (a) Original HR image (Goat). (b) LR observation (RMSE=19.16). (c) Bicubic interpolation (RMSE=16.74). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=15.85). (e) Reconstruction with the proposed Welsch type error norm (RMSE=15.37).

the level of detail in alignment. Note also that the mixing weights have been set to  $\alpha = 0.5, \beta = 0.25, \gamma = 0.25$  for both experiments.

In the first one, it has been assumed that accurate alignment is possible via enough numbers of landmarks. This case is especially valid in constrained image domains. Specifically, we have worked on face images in this experiment. Alignment of the reference image requires the identification of the landmarks. The general tendency is manually determining these features, or assuming that they are given. However,



**Figure 3.10:** Performance comparison of the Lorentzian and Welsch type M-estimators in the SRR scheme given in 3.13. (a) Original HR image (Airplane). (b) LR observation (RMSE=11.27). (c) Bicubic interpolation (RMSE=9.23). (d) Reconstruction by using Lorentzian norm in 3.13 (RMSE=5.18). (e) Reconstruction with the proposed Welsch type error norm (RMSE=5.24).



**Figure 3.11:** Behavior of the Welsch type evaluation function at different scales. (a) Original HR image. (b) LR observation obtained by 2x2 decimation, PSF blurring with a 5x5 Gaussian kernel  $N(0, 1)$  and additive white Gaussian noise with  $\sigma_n = 15$ . (c-h) Reconstruction results with (3.13) having different scale parameters  $c$  between 5 and 25.

these approaches are not practical for real-world applications, so we have considered a more practical approach by employing Active Appearance Models (AAM). Though



**Figure 3.12:** The mean-face used as the reference image while cloning the image details.

an introduction is given in Chapter 5, basically AAM is a modeling technique which also allows automatic extraction of the landmarks.

In order to increase the practicality, we have used the enhanced version of the mean-face, shown in Fig. 3.12, as the reference. Thus, we could skip the search process by using a common template for all test images. But, it should be noted that an individually selected reference image would provide better results than the common template.

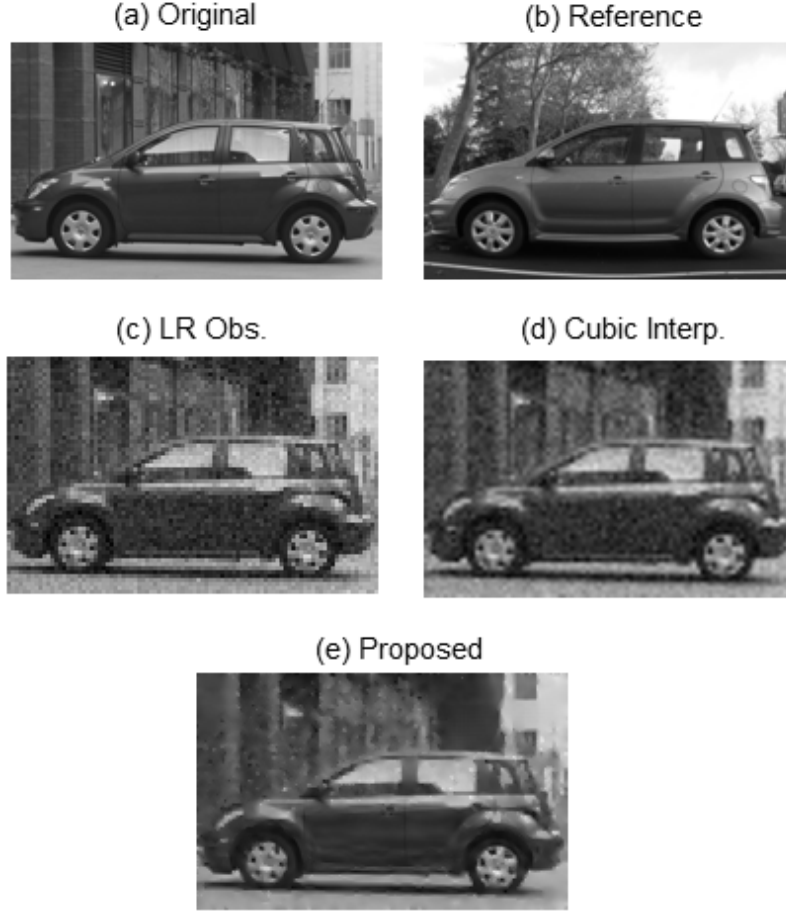
The results of this experiment are shown in Fig. 3.13. The reconstructions have much higher quality than the results of the traditional analytical methods. Especially, the image details, identifying the subject, could be reconstructed successfully.

After from this well-aligned case, we have experimented on the images which are not restricted to a specific domain; hence, they could only be aligned roughly. By using a limited number of landmarks, the reference image is warped to the LR observation. Though we have experimented with a wider set of images, each including different types of subjects, here we present the results on a car image shown in Fig. 3.14. This car image is interesting, since it is possible to observe the behavior of the proposed reconstruction scheme at different scene regions such as: the overlapping main region (car), the partially matching background, and the totally conflicting background (the tree in the reference).

As shown in Fig. 3.14, the proposed method has been able to clone the relevant details successfully while dismissing the unmatching ones.



**Figure 3.13:** Face reconstruction results using the proposed reconstruction scheme given in 3.11. (a) shows the original HR face images, (b) includes the LR observations obtained by  $2 \times 2$  decimation, psf blurring with  $N(0, 1)$  of size  $5 \times 5$  and additive white noise with  $\sigma_n = 10$ , (c) denotes the results of bicubic interpolation, and (d) consists of the reconstructions obtained by the proposed method.



**Figure 3.14:** Reconstruction of the car image by using the proposed method (3.11) and the reference image found from the repository search. (a) Original HR image. (b) The found reference. (c) LR observation obtained by 2x2 decimation, psf blurring with  $N(0, 1)$  of size 5x5 and additive white noise with  $\sigma_n^2 = 10$  (RMSE=19.16). (d) Reconstruction by bicubic interpolation (RMSE=16.28). (e) Reconstruction by the proposed method (RMSE=13.82).

### 3.4 Conclusion

We have introduced an SRR mechanism which does not require any prior training and works for the whole natural image space. Basically, we have utilized the anisotropic diffusion technique by using re-descending M-estimators. Specifically, the Welsch-type error norm has been employed both for smoothness and extrapolation. Compared to the other re-descending M-estimators, such as the popular Lorentzian norm, the Welsch norm shows better adaptation (better edge-stopping behavior) and provides significant computational conveniences. It has a closed form and can be differentiable everywhere. Especially in theoretical derivations and

mathematical programming, these computational advantages provide significant savings by eliminating the use of indicator functions of piecewise continuous functions. In addition, we have used a wealthier set of smoothing features rather than using only the first-order derivatives in  $x$  and  $y$  directions. Thus, we have relieved the blocking artifacts caused by the missing treatment of intermediate orientations and scales. Another significant contribution is utilizing a reference image for cloning the globally consistent realistic image details. A structurally and semantically close image have been searched from a repository (no need to be restricted to a particular domain) and used as the reference image while cloning the relevant image details. Structural similarity provides continuity in the result and semantic similarity helps incorporate realistic and reliable details. Since we cannot guarantee an exact alignment with the reference image, we have utilized the proposed robust form while incorporating the HF content. Thus, we could incorporate a significant amount of image details.





#### 4. LEARNING-BASED SUPER-RESOLUTION

In the previous chapter, we investigated the SRR problem with the cost function using robust error norms. As seen from this investigation, robust norms provide an adaptive modeling scheme, and this selective treatment fits well with the behaviors of the natural images. However, these functions mostly have non-convex or partially-convex structures, and numerical techniques are required in their optimization. Iterative solutions generally cause difficulties in online processing and are substituted by less adaptive but faster alternatives by sacrificing quality. As a remedy for such cases, in this section we have proposed an SRR solution, which is based on exploiting the enhanced Gaussian Conditional Random Field (GCRF). The selected modeling scheme provides the necessary adaptation for reconstruction quality without causing the aforementioned computational difficulties.

Traditional Gaussian Markov Random Field (GMRF) models are convenient to work with because they can be easily implemented using linear algebra routines. The inference can be especially accomplished quite efficiently through analytical expressions. However, this type of homogeneous models tend to over-smooth images and cause blurring. To overcome this problem, in his pioneering work [24] Tappen et al. has proposed a quite efficient alternative, called GCRF, by enhancing the traditional GMRF models in the form of conditional models. Moreover, in this enhanced GCRF model, the adaptation is increased more by employing custom weighting functions for local image regions. Hence, the required adaptation is obtained without sacrificing the computational conveniences of Gaussian models.

GCRF modeling has been used mostly for the decomposition of signals into their intrinsic components. For instance in [3], it has been employed for the denoising problem. The noisy input image is split into a clean image and a noise image. Moreover, in [24], a more generalized model has been used to obtain the albedo component of images. Considering the promising results on these analysis-type

problems, we have decided to investigate this theory for the solution of the SRR problem. But, SRR has a nature different from these analysis problems and requires also the reconstruction of the missing data. Therefore, we propose a mixed solution of analysis and synthesis schemes by using GCRF models.

The organization of this chapter is as follows: In Section 4.1 a brief introduction for the theoretical aspects of the GCRF model is given. Meanwhile the improvements proposed by Tappen et al. [24] are also described. Later in Section 4.2 we consider this GCRF model for the solution of the SRR problem. In Section 4.3, experimental results with the new approach are provided. The chapter is concluded with a discussion in Section 4.4.

#### 4.1 Definition

The role of the prior in regularization is the key for success in reconstruction. Since it is hard to model the whole natural image space analytically, building stochastic models, especially by exploiting local models, is more realistic. At that point, Markov Random Fields (MRF) provide us a powerful tool for doing this and are used as a common means for learning and inference on image models [9].

In MRF models, the relationships between neighboring nodes (also called cliques),  $I_c$ , are modeled by parametric local potential energies  $\mathbb{E}(I_c; \theta)$  in the form of Gibbs function  $f(I_c; \theta) = \exp(-\mathbb{E}(I_c; \theta))$ ; where  $I_c$  refers to the spatial neighborhood of pixels, and  $\theta$  is the set of parameters. Assuming that these local image regions are independent, the joint model for the image  $I$  is given as:

$$p(I) = \frac{1}{Z} \prod_{c=1}^C f_c(I_c; \theta_c) \quad (4.1)$$

where  $C$  is set of all cliques. In computer vision and image processing literature, there has been an intense interest in the representational power of the MRF modeling scheme, and a plethora of work has been published. One part of these studies is mainly interested in finding effective learning and inference algorithms on MRFs, such as [3, 84]. Although there are efficient algorithms (such as Graph-Cuts [85] and Loopy Belief Propagation [27]) for certain types of discrete valued graphs, learning and inference in MRFs are generally nontrivial problems. Another part of the research

focuses on defining better functions to denote the local potentials, such as [24, 22]. In this work, we mainly deal with defining an alternative potential energy function which also simplifies the learning and inference stages.

The general tendency in defining clique potentials is to use nonlinear and non-convex potential functions. As stated in the previous chapter, especially non-convex functions provide more adaptive models. For instance in the Field-of-Experts (FOE) model [22] Roth et al. has used Student-t distribution while defining the clique potentials as

$$\mathbb{E}(I_c) = \rho(I_c; \theta_c) \quad \text{where} \quad \rho(x; \gamma) = \left(1 + \frac{1}{2}(x)^2\right)^{-\gamma}. \quad (4.2)$$

Although these non-convex functions conform better with the non-uniform nature of the imaging space, both the parameter learning and inference in generic MRF models are quite difficult. Especially for learning, sophisticated sampling methods are required, and sampling algorithms are slow to converge.

On the other hand, Gaussian MRFs, where all the variables are jointly Gaussian, are particularly convenient to work with

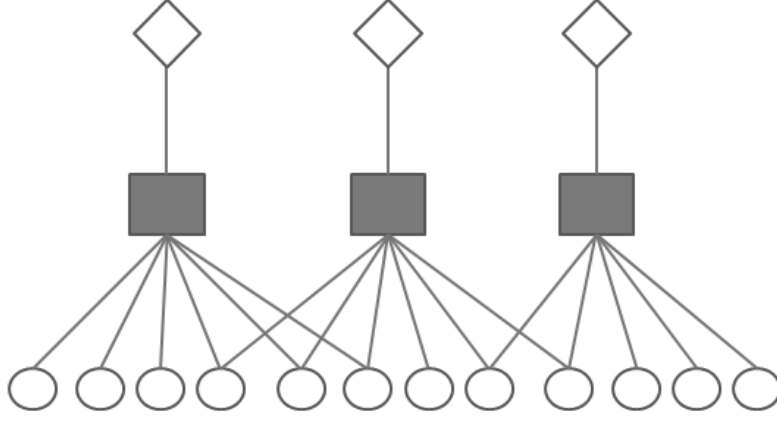
$$\mathbb{E}(I_c) = \rho(I_c) \quad \text{where} \quad \rho(x) = (x)^2. \quad (4.3)$$

Inference in Gaussian models can be easily accomplished using linear algebra. However, Gaussian MRFs can result in over-constrained images since the clique potentials are isotropic. But, it is possible to relieve this shortcoming by increasing the adaptation. One key attempt to increase the adaptation with Gaussian MRFs is to have the potential functions dependent on the measurement data  $O$  as

$$p(I|O) = \frac{1}{Z} \exp\left\{-\sum_{c=1}^C \mathbb{E}(I_c|O)\right\}. \quad (4.4)$$

These anisotropic models can overcome the weakness of the homogeneous ones by reconstructing piecewise constrained results with desirable properties. Since the potentials depend on the signal, these MRFs are no longer generative structures, but are instead conditional models. Therefore they are called Gaussian conditional random fields. Moreover, in [24] Tappen et al. generalizes this model to make it applicable for any purpose as

$$\mathbb{E}(I_c|O) = \sum_{(x,y) \in I_c} \sum_{i=1}^{N_f} ((I_c * \Gamma_i)(x,y) - r_i(x,y|O))^2, \quad (4.5)$$



**Figure 4.1:** Factor graph representation of the image model given in (4.6). Squares refer to factors, diamonds show the observation variables, and circles are the unknown variables representing the image.

where cliques are assumed as the square neighborhoods of each pixel  $(x,y)$ , and  $\Gamma_1 \dots \Gamma_{N_f}$  are the features designed as the convolution kernels characterizing the problem. The function  $r_i(x,y|O)$  is called the response estimator and refers to the expected or desired value of the convolution at that pixel,  $(I * \Gamma_i)(x,y)$ . For each feature  $\Gamma_i$ , the function  $r_i$  uses the observation  $O$  to estimate the value of the filter response. Considering all the cliques having potentials in the form of (4.5), the joint image model is defined as

$$p(I|O) = \frac{1}{Z} \exp \left( - \sum_{x,y} \sum_{i=1}^{N_f} ((I * \Gamma_i)(x,y) - r_i(x,y|O))^2 \right). \quad (4.6)$$

For more clarity the graphical representation of the model is given in Fig. 4.1.

Though, through this update the adaptation is increased for some cases, this GCRF model (4.6) still behaves uniformly in SRR case. To impose smoothness the features are selected as derivatives and their corresponding response estimators are set to 0 identically. As mentioned before, one way to avoid this over-smoothing is to use non-convex robust potential functions, such as the ones listed in Figures 3.2 and 3.3. Unfortunately, the convenience of the quadratic model is lost when these functions are used. Alternatively, the quadratic model can be improved by assigning weights to adjust the contribution of each potential. Incorporation of the weights can be expressed formally as

$$p(I|O) = \frac{1}{Z} \exp \left( - \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|O; \theta_i) ((I * \Gamma_i)(x,y) - r_i(x,y|O))^2 \right), \quad (4.7)$$

where  $w_i(x, y|O; \theta_i)$  are the positive weighting functions and  $\theta_i$  are their parameters. In [3], this model has been proposed for the denoising problem by imposing smoothness. For that purpose, the features  $\Gamma_i$  are selected as derivatives, and the response estimators  $r_i$  are set to 0 identically for all the features.

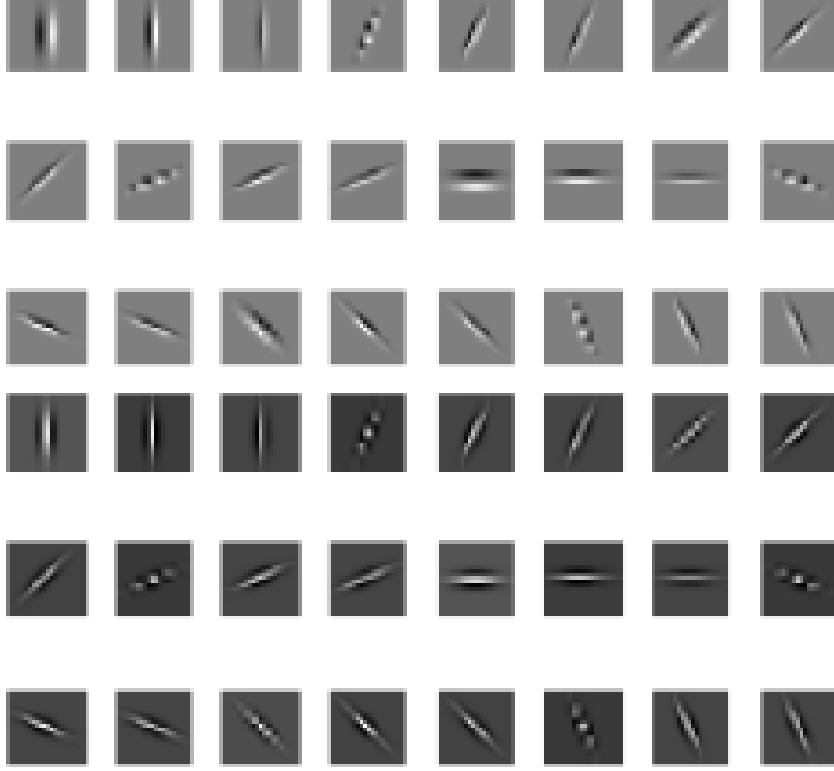
The weighting function  $w_i(x, y|O; \theta_i)$  can be designed in various forms. Except for being differentiable, there is no restriction. For instance in [3], Tappen et al. has suggested using the linear combination of a set of multi-scale oriented edge and bar filters

$$w_i(x, y|O; \theta_i) = \exp \left( \sum_{j=1}^{N_w} \theta_{ij} ((O * v_j)(x, y)) \right), \quad (4.8)$$

where  $v_j$  denote the filters and  $\theta_i = \theta_{i1}, \dots, \theta_{iN_w}$  are the regression coefficients. The exponential here ensures that the weight is always positive. Tappen et al. [3] uses this function in his denoising solution to guess where the edges occur in the image and to reduce the smoothness constraints appropriately. Sensitivity to extended edges has been increased by using weight function filters  $v$  operating at multiple scales (having the sizes of 11 pixels, 21 pixels and 31 pixels). Moreover, a third set of responses has been additionally created by adding the squared responses of corresponding edge (the first three rows in Fig. 4.2) and bar filters (the remaining 24 in Fig. 4.2).

The parameter set of a GCRF model of this kind consists of the regression coefficients used in the weight functions. That means the total number of parameters to be estimated is  $N_f x N_w$ , and even in a moderate setup this number can be large enough (e.g. when 6 derivative features  $\Gamma_i$  and 72 weight function filters  $v_k$ , as in the above example, are used, then the number of parameters will be 432). Using a small number of features generalizes the model too much, whereas using lots of features would fit the training data, especially when the training data is limited. Hence, the total number of features is critical and a good balance should be kept.

As stated in [3], the GCRF model, given in (4.7), can be motivated in two ways; probabilistically as a CRF model or as an estimator based on the minimization of a cost function. Though we consider the probabilistic interpretation of the model while describing our solution in Section 4.2, we also give the relation with the cost function perspective. As seen later, with the appropriate selection of the solution components ( $w, r$ , and  $f$ ), the learning can be greatly simplified. Therefore, in the next



**Figure 4.2:** Weighting function features used in [3] for denoising images. The first three rows show the edge filters and the remaining are bar filters.

two subsections, we introduce both the inference and learning issues of the enhanced GCRF model (4.7) from both interpretations by mostly following the notations used in [3].

#### 4.1.1 Inference

The clique potentials, the exponent in (4.7), can be written in matrix form by creating a set of matrices  $F = \{\Gamma_1 \dots \Gamma_{N_f}\}$ . Each matrix  $\Gamma_i$  performs the same set of linear operations by convolving an image with a filter  $\Gamma_i$ . In other words, if  $I(x, y)$  is the two dimensional image representation and  $I$  is the vectorial lexicographical ordering, then  $\Gamma_i I$  is identical to the convolution  $(I * \Gamma_i)(x, y)$  at all pixels,  $\forall (x, y) \in I$ . These matrices can then be stacked and the exponent can be rewritten as

$$\sum_{x,y} \sum_i^{N_f} w_i(x, y | O; \theta_i) ((I * \Gamma_i)(x, y) - r_i(x, y | O))^2 \approx (FI - R)^T W(O; \theta) (FI - R), \quad (4.9)$$

where

$$F = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_{N_f} \end{bmatrix}, \quad R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{N_f} \end{bmatrix}, \quad (4.10)$$

$$W(O; \theta) = \begin{bmatrix} W_1(O; \theta_1) & \cdot & \cdots & \cdot \\ \cdot & W_2(O; \theta_2) & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & W_{N_f}(O; \theta_{N_f}) \end{bmatrix}.$$

The block-diagonal matrix  $W(O; \theta)$  is a function of the observation  $O$  and the parameters  $\theta$ . Each element along the diagonal of  $W_i(O; \theta_i)$  matrices is equal to the weight of a term at a particular locality,  $w_i(x, y|O; \theta_i)$ .

The similarity of (4.9) with the exponent of a multivariate normal distribution,  $(I - \mu)^T \Sigma^{-1} (I - \mu)$ , allows us to re-write (4.9) in the form of a normal distribution having the parameters

$$\mu = (F^T W(O, \theta) F)^{-1} F^T W(O, \theta) R \quad \Sigma^{-1} = F^T W(O, \theta) F. \quad (4.11)$$

Notice that the difference between (4.9) and the exponent of  $N(I; \mu, \Sigma)$  is constant and equal to  $R^T R - \mu^T \mu$ . However, since it is the same for all  $I$ , it only affects the normalization constant,  $Z$ . The relative probabilities do not change.

As stated before, the GCRF model can also be motivated from a cost function point of view. Let  $h(O; \theta)$  be an estimator using the observation  $O$  to estimate an image. The estimate  $\hat{I}$  is the image that minimizes the quadratic cost function

$$C(I|O; \theta) = \sum_{x,y} \sum_i^{N_f} w_i(x, y|O; \theta_i) ((I * \Gamma_i)(x, y) - r_i(x, y|O))^2. \quad (4.12)$$

The minimum of this quadratic structure can be computed via pseudo-inverse. Using the matrix notation of (4.12), the inverse can be expressed as

$$h(O; \theta) = (F^T W(O; \theta) F)^{-1} F^T W(O, \theta) R. \quad (4.13)$$

As seen from (4.11) and (4.13), the mode of the joint distribution (namely the mean of the MAP estimator given in (4.9)) is equal to the minimum of the quadratic cost expression of (4.12). Then, it can be stated that the solution is the result of the

following analytical expression

$$\hat{I} \cong \arg \max_I p(I|O) \cong (F^T W(O; \theta) F)^{-1} F^T W(O, \theta) R. \quad (4.14)$$

#### 4.1.2 Learning

Assuming that we already have the features  $\Gamma_i$  and response estimators  $r_i$ , the remaining unknowns in (4.14) are the parameters of the weighting functions,  $\theta_i$ . Traditionally the parameters of the Conditional Random Fields (CRF) are found by maximizing the likelihood of the training data, which is known as the Maximum Likelihood Estimation (MLE) [9].

Considering the GCRF model given in (4.9), the log-likelihood of a training sample  $T$  under the condition of an associated observation  $O_T$  is denoted as

$$LL(T|O_T) = - \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|O_T; \theta_i) ((T * \Gamma_i)(x,y) - r_i(x,y|O_T))^2 \quad (4.15)$$

$$- \log \int \exp \left( - \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|O_T; \theta_i) ((I * \Gamma_i)(x,y) - r_i(x,y|O_T))^2 \right) dI.$$

Then the partial derivative with respect to the parameter  $\theta_i$  will be

$$\frac{\partial LL(T|O_T)}{\partial \theta_{ij}} = - \sum_{x,y} \sum_{i=1}^{N_f} \frac{\partial w_i(x,y|O_T; \theta_i)}{\partial \theta_{ij}} ((T * \Gamma_i)(x,y) - r_i(x,y|O_T))^2 \quad (4.16)$$

$$+ \frac{1}{Z} \int \sum_{x,y} \sum_{i=1}^{N_f} \frac{\partial w_i(x,y|O_T; \theta_i)}{\partial \theta_{ij}} ((I * \Gamma_i)(x,y) - r_i(x,y|O_T))^2$$

$$\exp \left( - \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|O_T; \theta_i) ((I * \Gamma_i)(x,y) - r_i(x,y|O_T))^2 \right) dI.$$

Notice that the integration term of (4.16) can be rewritten as the expected value (**ExpVal**[]) of the total energy function

$$\mathbf{ExpVal}[\mathbb{E}(I|O_T)] = \frac{1}{Z} \int \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|O_T; \theta_i) ((I * \Gamma_i)(x,y) - r_i(x,y|O_T))^2 \quad (4.17)$$

$$\exp \left( - \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|O_T; \theta_i) ((I * \Gamma_i)(x,y) - r_i(x,y|O_T))^2 \right) dI.$$

Though it is not clearly stated in [3], the integral in (4.16) differs from that of (4.17) with the form of the weighting term. In (4.16), the derivative of the weighting function



is used differently. However, we can easily keep the equivalence by multiplying the integral term in the gradient of the likelihood function (4.16) with a block-diagonal matrix  $\alpha$  consisting of constant  $\frac{w_i(O_T, \theta_i)}{\partial w_i(O_T, \theta_i) / \partial \theta_{ij}}$  values. Recall also that in (4.7) the potential functions have a Gaussian form. This means that the expectation in (4.17) is equal to the mean  $\mu$  defined in (4.11). Replacing the integral with the mean and using the matrix forms of the operators, (4.16) can be re-written as

$$\frac{\partial LL(T|O_T)}{\partial \theta_{ij}} = -(FT - R)^T \frac{\partial W(O_T; \theta)}{\partial \theta_i} (FT - R) + \alpha (F^T W(O_T; \theta) F)^{-1} F^T W(O_T; \theta) R \quad (4.18)$$

As stated in [3], it is also possible to learn these parameters by following a discriminative learning strategy. The penalty is expressed using a loss function  $L(\hat{I}, T)$  that assigns a loss for the intermediate estimate  $\hat{I}$  based on its distance from the ground-truth image  $T$ . It is assumed that  $L$  is designed as the squared difference:  $L(\hat{I}, T) = (\hat{I} - T)^T (\hat{I} - T)$ . In the GCRF model, the mode of the conditional distribution is the conditional mean, and thus the cost  $C(T|O_T; \theta)$  associated with a particular set of parameters  $\theta$  is

$$C(T|O_T; \theta) = L\left((F^T W(O_T; \theta) F)^{-1} F^T W(O_T; \theta) R, T\right). \quad (4.19)$$

The parameters  $\theta$  can be found by minimizing  $C(T|O_T; \theta)$ . The optimization is performed through a gradient descent technique, where the gradient is found via

$$\frac{\partial C(T|O_T; \theta)}{\partial \theta_{ij}} = 2\left((F^T W(O_T; \theta) F)^{-1} F^T W(O_T; \theta) R - T\right). \quad (4.20)$$

$$\begin{aligned} & \left( (F^T W(O_T; \theta) F)^{-1} F^T \frac{\partial W(O_T; \theta)}{\partial \theta_{ij}} \right). \\ & \left( -F (F^T W(O_T; \theta) F)^{-1} F^T W(O_T; \theta) R + R \right). \end{aligned} \quad (4.21)$$

Although the system (4.9) is presumably quadratic and linear systems occur, as seen above, both of the learning strategies do not result in estimators having explicit forms. Therefore, numerical optimization techniques are required to find the MLE of the parameters.

Despite this similarity, we have provided both alternatives for learning. For the appropriate selections of weighting  $w_i$  functions, the estimators are greatly simplified and allow analytical expressions. Thus, the learning can also be possible for large-scale

images. As seen from (4.18) and (4.21), the most expensive step in learning is computing the inverse of the weighting functions as part of the covariance. The complexity of computing the gradient is  $O(n^3)$ , and these heavy computational implications may cause serious limitations with large-scale images.

## 4.2 GCRF Devoted To SRR

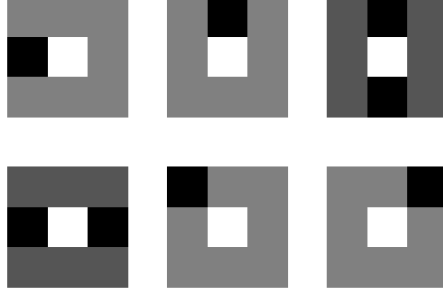
A typical GCRF model (4.9) requires the design of three main components; weights ( $w_i$ ), features ( $\Gamma_i$ ), and response estimators ( $r_i$ ); so as to maximize the adaptation to the needs. Each of the components can be designed independently corresponding to the different characteristics of the problem.

In literature, traditional applications of the GCRF model have always been related with the decomposition of mixed signals, because its flexible form fits well with analysis purposes. For instance in [24], it has been used for the identification of the albedo component of the input image. Moreover, in [3], the GCRF model has been suggested for the denoising problem by using the following posterior probability

$$p(I|O) = \frac{1}{Z} \exp \left( -\|HI - O\|_2^2 - \sum_{x,y} \sum_{i=2}^{N_f} w_i(x,y|O;\theta_i) ((I * \Gamma_i)(x,y))^2 \right), \quad (4.22)$$

where  $I$  is the denoised image and  $O$  is the noisy observation. Also, the solution components,  $\Gamma_i, w_i, r_i$ , are selected as in the basic setup (it was described before as:  $\Gamma_i$  consisting of derivatives given in Fig. 4.2,  $r_i$  are all set to 0, and  $w_i$  are determined with the regression expression (4.8) utilizing the weighting filters shown in Fig. 4.3). In (4.22), to keep the system invertible, the first feature  $\Gamma_1$  is selected as the filter corresponding to the PSF with  $r_1 = O$  and  $w_1 = 1$ . In other words, the first feature constraints the solution to be somewhat close to the observation under the assumed deformation model. This selection can be interpreted as the Bayesian interpretation of the GCRF model, where these initial components correspond to the likelihood term.

However, SRR is different from this type of analysis problem and require the synthesis of the missing image details completely lost during image formation. We have suggested using a reference data source to clone the relevant image details. By the addition of this cloning term, the complete reconstruction scheme of the HR image  $I_H$  given the LR observation  $I_L$  can be expressed through the following posterior



**Figure 4.3:** Derivative features used to impose smoothness in the solution.

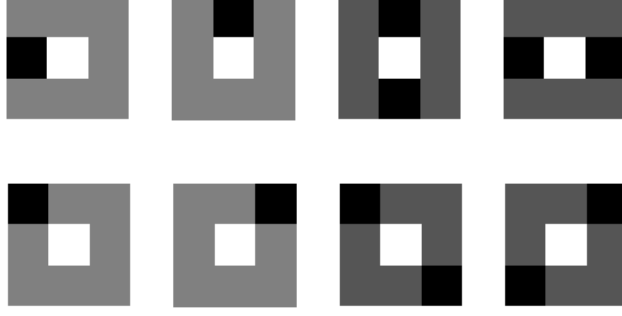
distribution

$$p(I_H|I_L) = \frac{1}{Z} \exp \left( \begin{array}{l} -\sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|I_L; \theta_i) ((I_H * \Gamma_i)(x,y) - r_i(x,y|I_L))^2 \\ -\sum_{x,y} \sum_{j=1}^{N_s} w_j^s(x,y|S, I_L; \theta_j) ((I_H * \Gamma_j^s)(x,y) - r_j^s(x,y|S))^2 \end{array} \right), \quad (4.23)$$

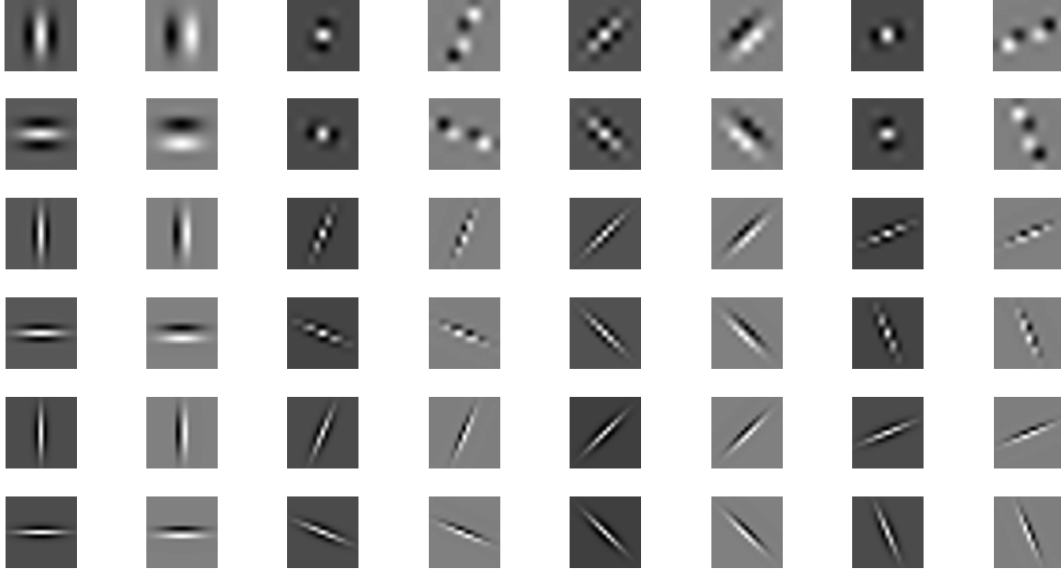
where  $S$  refers to the outside data,  $\Gamma_j^s$  are the features extracting the relevant details (mostly corresponding to high-frequency components),  $r_j^s$  are the response estimators providing desired content, and the  $w_j^s$  are the evaluation terms which obstruct copying irrelevant details. Thus, the new reconstruction scheme consists of three more components ( $w^s$ ,  $f^s$  and  $r^s$ ) in addition to  $w$ ,  $f$  and  $r$  of (4.9). In the rest of this section we provide the design details for these 6 components.

- **Image features:** The most generic regularity known for the natural image space is being piece-wise smooth [23]. So as to impose smoothness, we suppress excessive gradients by using derivative features in various orders. As explained in Chapter 2, there are some basic principles to be followed while designing these feature sets. Though the variety and number of features are critical for better analysis, we should also consider the computational burden which grows with every additional feature. For smoothness and cloning parts we use different feature sets. In Fig. 4.3, the derivative features used to impose smoothness are shown. Similarly, for the cloning term we again use derivatives, as shown in Fig. 4.4, but this time the number of features is increased to capture more details.
- **Weighting Functions:** The single restriction on the weight functions is that they be differentiable. For the smoothing term we employ a slightly different form of the regression expression suggested by Tappen [3]:

$$w_i(x,y|I_L; \theta_i) = \exp \left( \sum_{k=1}^{N_w} \theta_{ik} ((\uparrow I_L * v_k)(x,y)) \right) \quad (4.24)$$



**Figure 4.4:** Derivative features used to clone image details through a reference image.



**Figure 4.5:** Elongated edge and bar filters used to build the weighting function in the form of a regression.

where  $\uparrow$  is the up-sampling operator. On the other hand, to clone the relevant image details and penalize the mismatching ones, we propose the following weighting function:

$$w_j^s(x, y | S, I_L; \theta_j) = \exp \left( \sum_{k=1}^{N_w} \theta_{jk} |(\uparrow I_L * v_k)(x, y) - (S * v_k)(x, y)| \right). \quad (4.25)$$

For both weight functions we use a similar set of band-pass features consisting of elongated edge and bar filters, as shown in Fig. 4.5. Different from the previous approaches [24, 3], we try to keep the number of features limited so as to avoid computational difficulties, such as data fitting and complexity. Also, we adjust the scales of these filters larger since we need to convolve with the interpolated images as different than the previous analysis applications.

- **Response Estimators:** Traditionally, the response estimators are conditioned on the observation and interpreted as the expected behavior at that location or clique. For instance, to impose smoothness each response estimate is set to zero. We also use similar type response estimators for the smoothness term

$$r_i(x, y | I_L) = 0, \quad \text{for } \forall i, \quad \text{and } \forall (x, y) \in I_H, \quad (4.26)$$

However considering only the expectations will not be enough for SRR. So, to synthesize the missing data, the response estimators should also provide the desired information. In other words, it is expected that the response estimator is capable of cloning the relevant image details from a trusted data source. Stated in the previous chapters, we have used semantically close images as the data source,  $S$ . An HR match of the observation is found from a repository, and by using this reference image  $S$  we design the response estimator for the data cloning term as

$$r_j(x, y | S) = (\Gamma_j^S * S)(x, y), \quad (4.27)$$

where the  $\Gamma_j^S$  are the custom feature extractors used to identify edges, lines, and any useful object features in the scene.

It is also possible to find alternative data source designs in literature. A review has been provided in Chapter 2, where the dictionary based approaches (e.g. [24]) are the most popular ones. However, these approaches are mostly based on local models and suffer from discontinuity artifacts and heavy computations.

By using the posterior probability distribution (4.23), the reconstruction is defined as the inference

$$\hat{I}_H = \arg \max_{I_H} p(I_H | I_L). \quad (4.28)$$

The parameters of (4.23) are learned through the discriminative learning described in Section 4.1.2 and the inference can be obtained analytically as shown in Section (4.1.1).

### 4.3 Experiments

In Sections 4.1 and 4.2, we provided the necessary learning and inference expressions for the proposed reconstruction scheme. It was shown that the reconstruction can be obtained quite efficiently through an analytical expression. In addition to such computational advantages, in this section we investigate the performance of the proposed solution in reconstruction quality through a set of experiments.

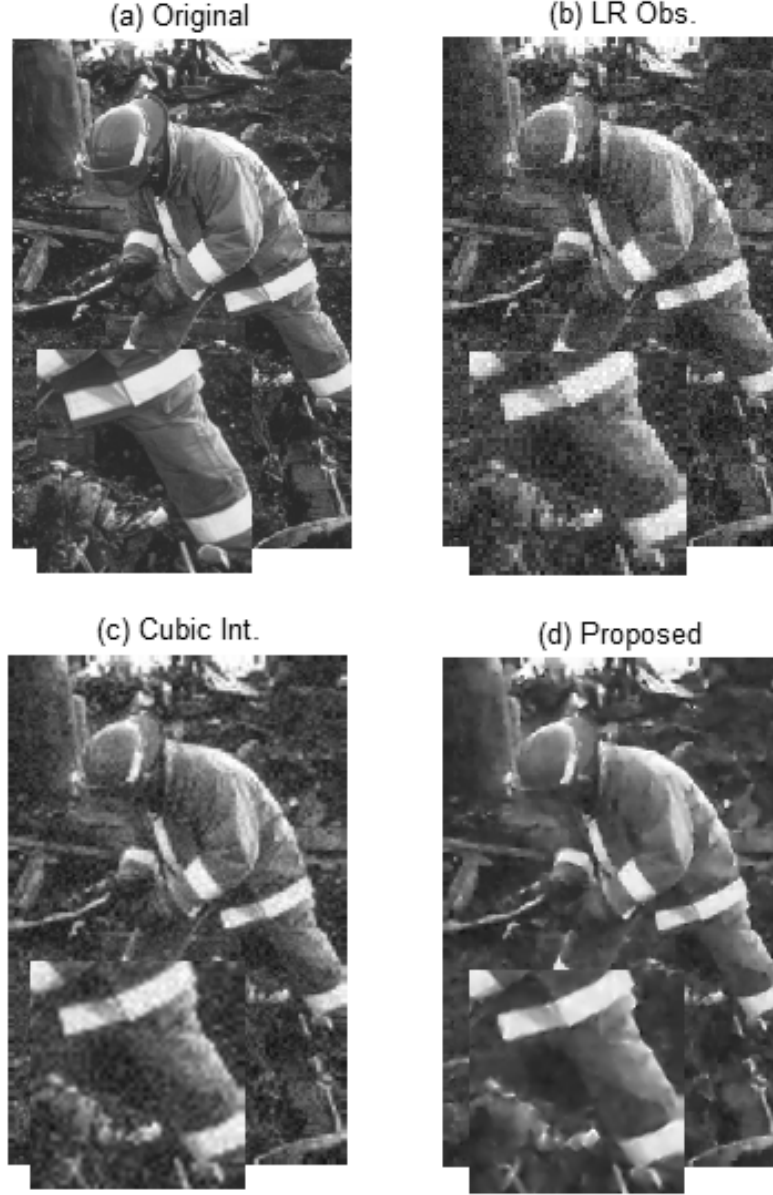
First, we have considered the simplified reconstruction scheme where no reference data sources are used. Thus, we have aimed for observing the adaptation power of the proposed GCRF modeling scheme. At that point, the posterior distribution, given in (4.23), turns to

$$p(I_H|I_L) = \frac{1}{Z} \exp \left( - \sum_{x,y} \sum_{i=1}^{N_f} w_i(x,y|I_L; \theta_i) ((I_H * \Gamma_i)(x,y) - r_i(x,y|I_L))^2 \right). \quad (4.29)$$

In this expression the components have been selected as follows; the smoothness features  $\Gamma_i$  are the derivatives of the first 2 order at 4 orientations, the weighting function is designed as in (4.24), the regression features  $v_k$  are edge and bar filters shown in Fig. 4.5, and the response estimators  $r_i$  are all 0 for smoothness. The observations have been obtained by 2x2 decimation, blurring with a 5x5 symmetric smoothing kernel having the parameters  $N(0,1)$ , and corrupting with additive white noise with  $\sigma = 10$ . Three test images have been deformed according to this formation model, and their reconstructions have been compared with the bicubic interpolation as shown in Figures 4.6-4.8.

The results in Figures 4.6-4.8 clearly show that the weighting function (4.24) used in the GCRF prior can successfully identify the edge regions and ignores the smoothness constraints at those parts. Thus, piece-wise smooth reconstructions can be obtained, while the uniform kernel interpolation [34, 35] makes the images excessively blurry.

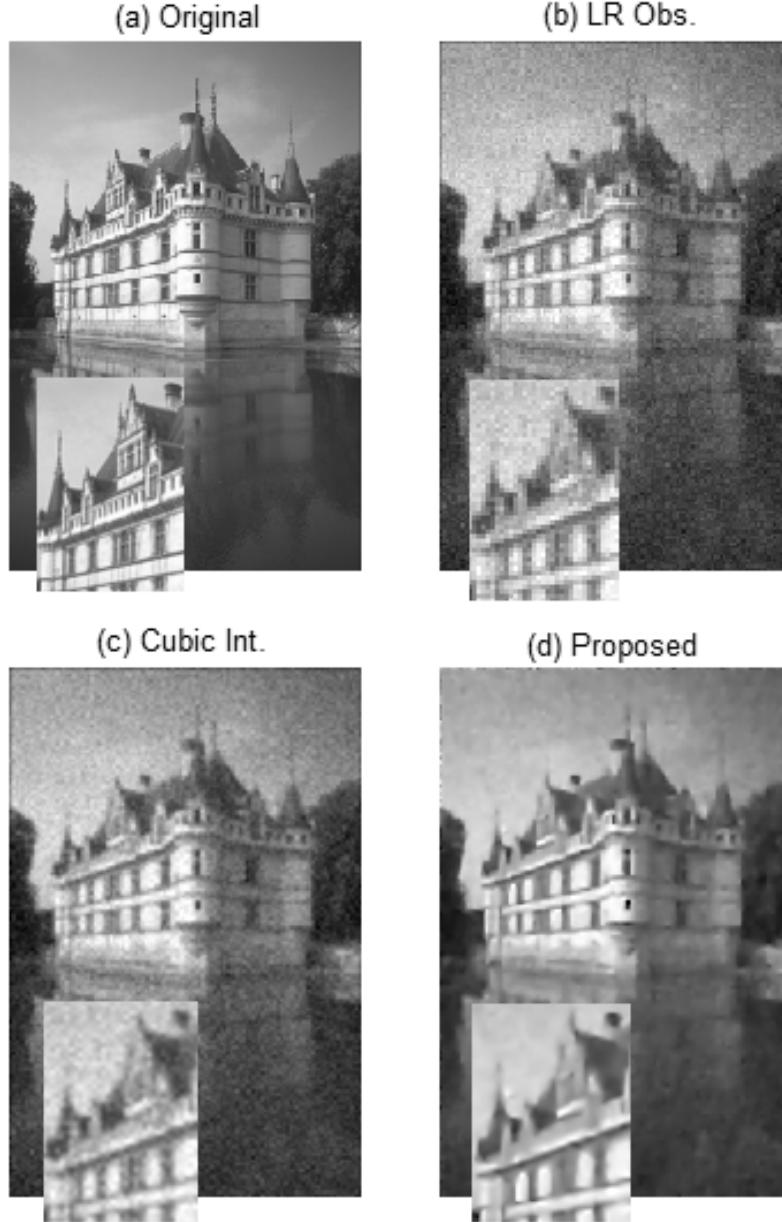
Later, we have experimented with our proposed reconstruction scheme (4.23) by using reference data. As we also did in the previous chapter, structurally and semantically close images, found from some repository by using the observation, have been employed as the data source. To maximize the structural similarity, the found reference has been initially aligned with the observation. Depending on the available resources



**Figure 4.6:** Reconstruction performance of the GCRF image prior where only the piece-wise smoothness is imposed. (a) Original HR image (Fireman). (b) LR observation (RMSE=20.58). (c) Reconstruction by bicubic interpolation (RMSE=17.79). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=15.98).

and the problem setup, the detail level of the alignment may change. In the experiments this variety has been included through the cases having high and low accuracies in alignment.

Especially in constrained image domains, such as face, registration is easier and the alignment accuracy is higher. The general tendency in applications working on such constrained image domains is to determine the registration parameters automatically by inferring from learned models. In this experiment, we have used samples which are

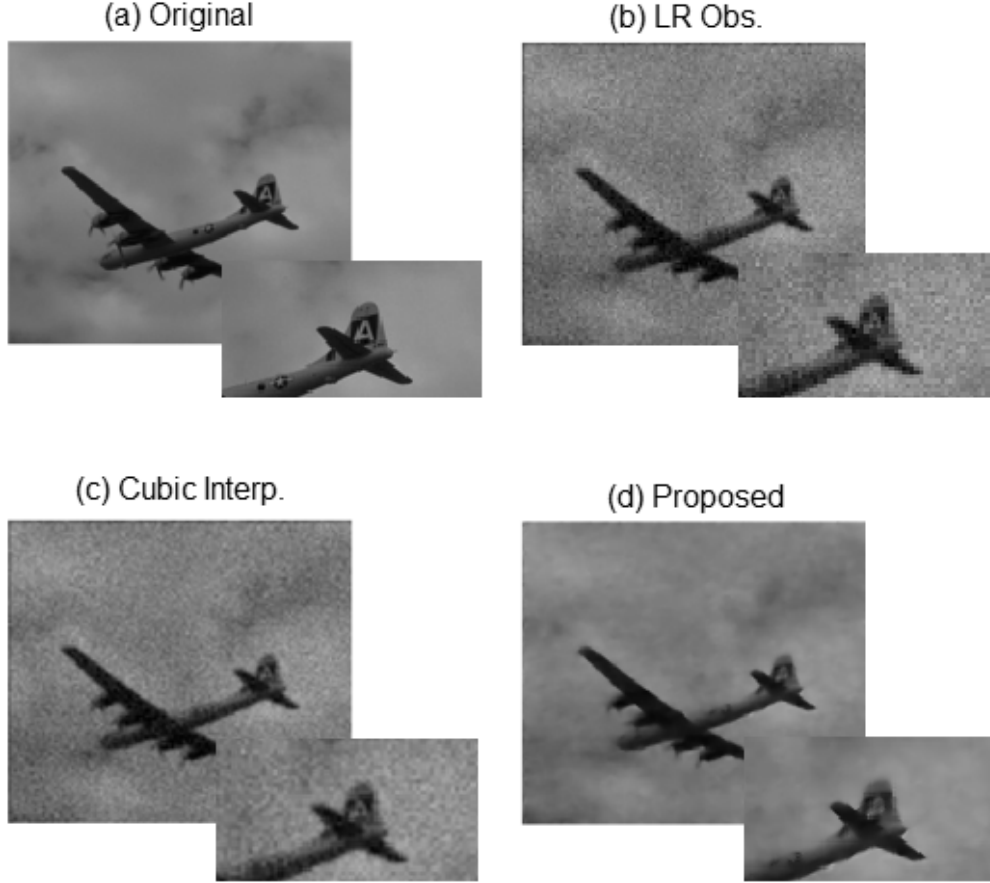


**Figure 4.7:** Reconstruction performance of the GCRF image prior where only the piece-wise smoothness is imposed. (a) Original HR image (Castle). (b) LR observation (RMSE=18.65). (c) Reconstruction by bicubic interpolation (RMSE=16.56). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=13.25).

automatically aligned onto the mean shape by using AAM (more detail of this process is given in Chapter 5). In Fig. 4.9 we compare the reconstruction results with the bicubic interpolation.

As seen in Fig. 4.9, the reconstruction is much better than the classical kernel interpolation. A significant amount of image details could be added while successfully rejecting the irrelevant ones. To show the contribution of the cloning more clearly, the



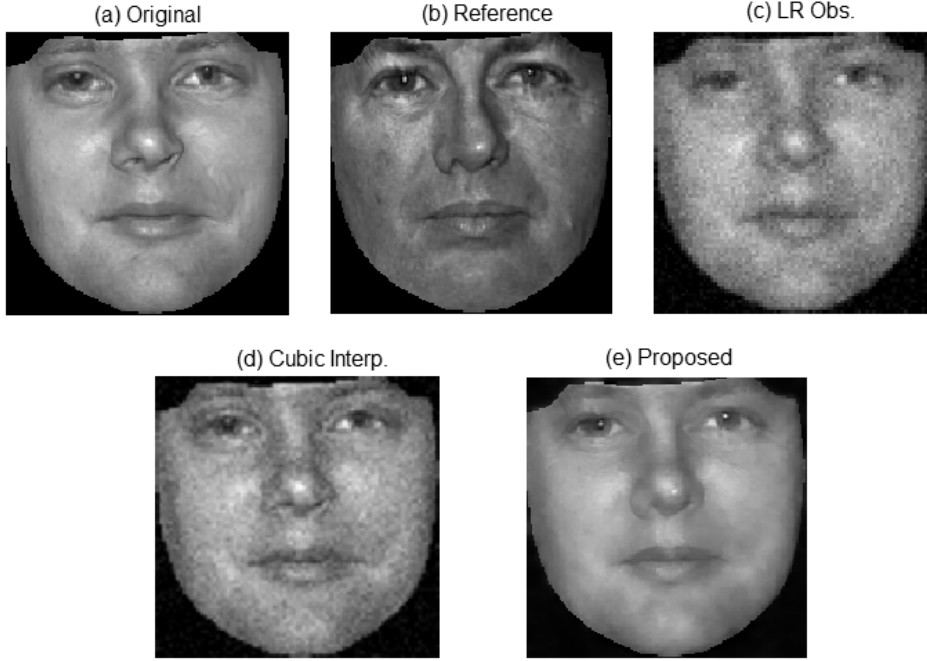


**Figure 4.8:** Reconstruction performance of the GCRF image prior where only the piece-wise smoothness is imposed. (a) Original HR image (Airplane). (b) LR observation (RMSE=11.27). (c) Reconstruction by bicubic interpolation (RMSE=9.23). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=4.41).

same experiment in Fig. 4.9 has been repeated without using additional data and the results shown in in Fig. 4.10.

In the results of Fig. 4.10, the smoothness constraints have become prominent in the reconstruction. The piece-wise smooth result includes less image details than the reconstruction shown in Fig. 4.9. Despite this decline in the perception, the RMSE value of the reconstruction in Fig. 4.10 with the proposed method is slightly lower than the one in Fig. 4.9. This is mainly caused by the mismatching faint details cloned. The adaptation is gained through learning and some little gap in this adaptation should have been accepted initially as in this case.

Then, we have considered the other case where the images can be aligned just roughly. In this scenario, the imaging space is not restricted as in the previous case and may consist of complex textures. The single restriction is that the scene consists of

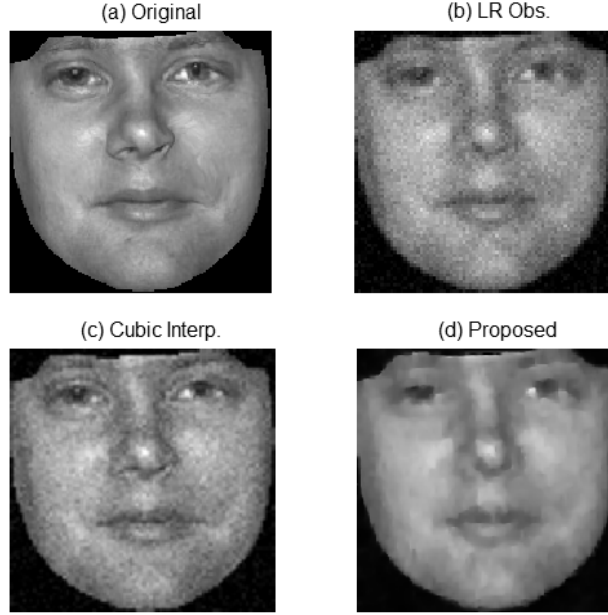


**Figure 4.9:** Reconstruction performance of the GCRF image prior where not only the piece-wise smoothness is considered but also the data cloning constraints are incorporated. (a) Original HR face. (b) Reference HR image which is correctly aligned with the observation. (c) LR observation (RMSE=14.61). (d) Reconstruction by bicubic interpolation (RMSE=12.46). (e) Reconstruction by inferring from the posterior given in (4.23) (RMSE=10.77).

object/objects having enough discrimination to be used in a reference search. We have considered the following car images for this experiment. As shown in Fig. 4.11, the objects are big enough for identification, though the background scenes are different. Before the reconstruction we have again aligned the found reference image with the observation roughly by using the main car parts (two wheels, one headlight and one stop-light), which are easily detectable in both the observation and the reference. The LR input has been obtained by using the same parameters of the observation model described above. The results are compared with the cubic interpolation in Fig. 4.11.

Although the reference has been aligned roughly, as shown in Fig. 4.11, a significant amount of details could be added. The contribution is more realizable when compared with the results in Fig. 4.12 where the same experiment has been repeated without using the cloning constraints.

Due to the selective treatment and the additional data cloning constraints, the method shows robustness against noise. In Fig. 4.13, we provide the reconstruction results



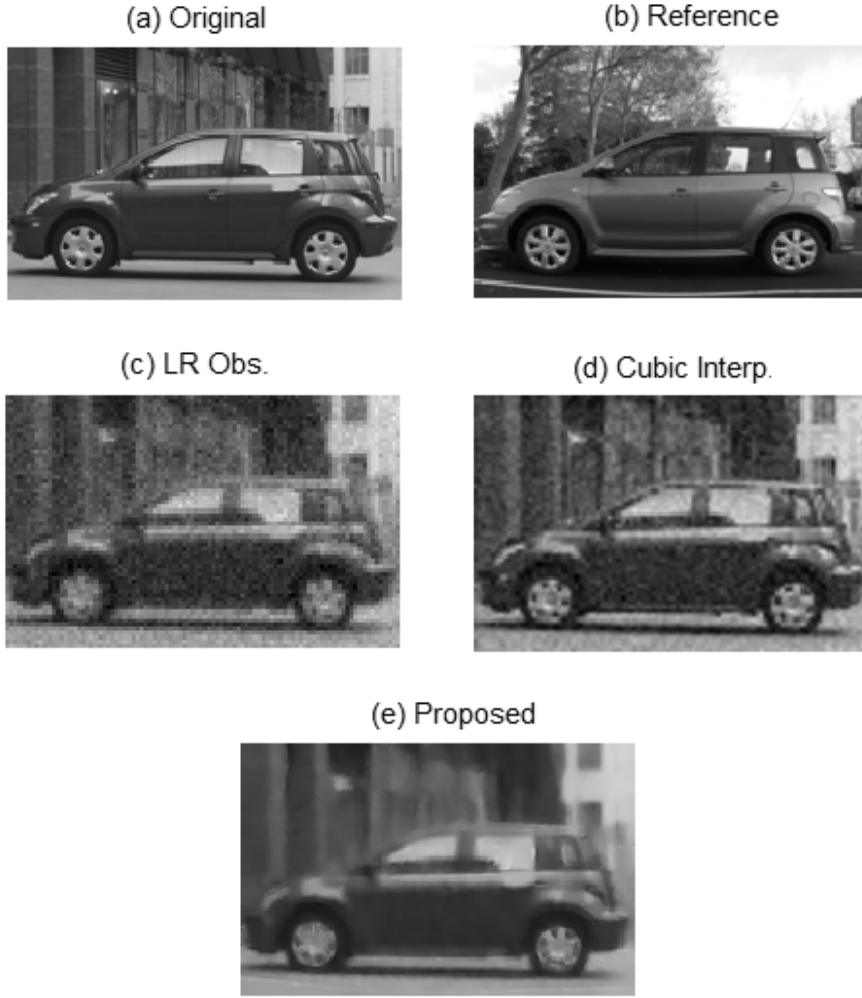
**Figure 4.10:** Reconstruction quality degrades when the data cloning constraints are neglected. (a) Original HR face. (b) LR observation (RMSE=14.62). (c) Reconstruction by bicubic interpolation (RMSE=12.46). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=10.71).

under noises in varying severity and compare the results again with the linear interpolation results.

#### 4.4 Conclusion

In this section, we have considered the SRR in cases where the fastest online reconstruction is strongly desired and enough data resource is available for any offline learning process. We have approached the problem from the statistical perspective and proposed a solution based on MAP estimation. The solution space has been represented by defining the posterior distribution in the form of an enhanced GCRF, which is in fact parametrically weighted GCRF and initially proposed by Tappen et al. in [3]. In addition to the better representational power, the used GCRF modeling scheme provides significant computational conveniences, such as:

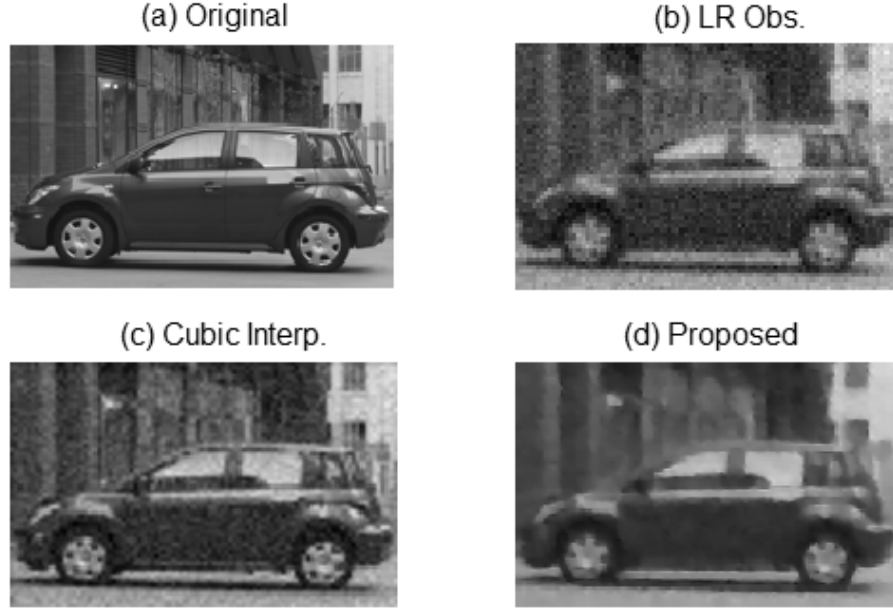
- **Computational cost:** GCRF models do not rely on machinery from convex optimization. In learning stage, the gradient can be calculated analytically. Even, when the images are used in appropriate sizes, the inference can also be performed analytically. Compared to the non-convex MRF schemes, which need complex



**Figure 4.11:** Reconstruction performance of the GCRF image prior where not only the piece-wise smoothness is considered but also the data cloning constraints are incorporated. (a) Original HR image. (b) Reference HR image which could be aligned with the observation roughly. (c) LR observation (RMSE=19.22). (d) Reconstruction by bicubic interpolation (RMSE=16.31). (e) Reconstruction by inferring from the posterior given in (4.23) (RMSE=13.39).








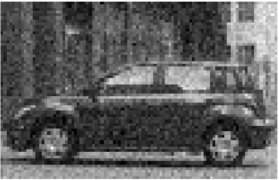


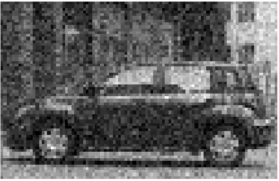


sampling algorithms and numerical optimization techniques, having such analytical structures both in learning and inference make this method advantageous.

- **Flexibility:** The proposed GCRF model allows excessive customization (namely adaptation) through various components (weights, features and the response estimators). Thus wide-variety of *a priori* information can be incorporated into the solution. However this increased adaptation comes with a cost in training since many parameters are to be learned.



**Figure 4.12:** Reconstruction quality degrades when the data cloning constraints are neglected. (a) Original HR image. (b) LR observation (RMSE=19.20). (c) Reconstruction by bicubic interpolation (RMSE=16.27). (d) Reconstruction by inferring from the posterior given in (4.29) (RMSE=13.32).

In addition to these, we have suggested using data cloning constraints. These additional constraints have been incorporated into the posterior through custom weighting functions and response estimators. The comparative experiments prove that the proposed solution is quite successful in cloning the relevant image details while dismissing the unmatched ones.

Original			
			
Noise Var.	LR Obs.	Cubic Interp.	Proposed
0	 RMSE=16.42	 RMSE=14.18	 RMSE=12.14
10	 RMSE=19.14	 RMSE=16.23	 RMSE=13.39
15	 RMSE=22.18	 RMSE=18.59	 RMSE=14.15
20	 RMSE=25.58	 RMSE=21.30	 RMSE=15.56

**Figure 4.13:** Reconstruction performance under different noise levels.

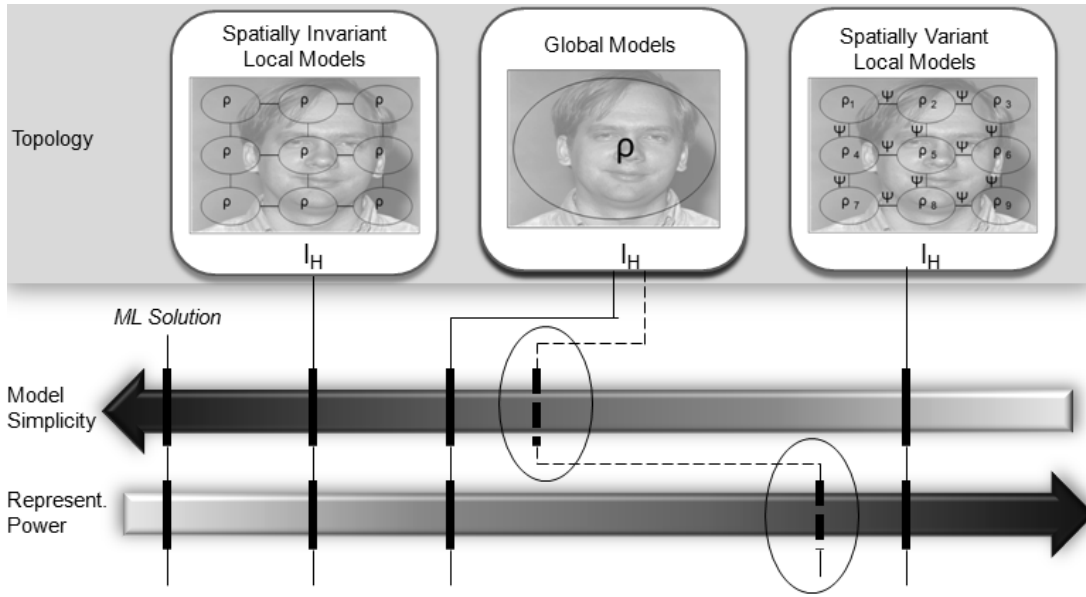
## 5. FACE HALLUCINATION

In this chapter we particularly consider a restricted case of the SRR where again the fastest online reconstruction is highly desired and sufficient resources are available for offline learning. Different from the previous cases, specifically the problem setup assumes that the images used in learning are highly correlated with the HR image to be estimated. That means the imaging space is constrained to some specific scenes such as face, plate, bone, cell.

Although the proposed solution is valid for any type of rigid object scenes, due to its wide-variety of applications, here we are only interested in face images. The exact problem can be defined as: given a single LR face image, infer the much higher resolution of it under the known deformation model. This constrained version of the generic super-resolution problem is specifically known as Face Hallucination [10] in literature. The solution corresponds to the inverse of the image formation, but the backward model has an ill-posed nature as mentioned before. Therefore, the reconstruction is defined as an estimation in the form of Maximum A-Posteriori (MAP).

In literature, the Maximum Likelihood (ML) solution is defined almost the same as the data fidelity constraint [13]. The ML estimate mainly contributes to the global characteristics of the image space, because the low-frequency content of the original HR image is encoded in the LR observation [5]. However higher frequencies are critically important in successive applications, such as face recognition and tracking, where the resolution is normally quite low but important for identification. These applications require the reconstruction of these missing image details. At that point, *a priori* information plays an important role in adding relevant details, which correspond to the lost mid/high frequency content.

In opposition to the ML solution, various prior designs [8] have been proposed, as reviewed in Chapter 2. These priors vary from spatially invariant (homogeneous)



**Figure 5.1:** Comparison of image prior models in terms of representational power and computational complexity. Continuous lines show the corresponding behavior for different topologies, and the dash line is the target behavior of this work. The topmost row shows roughly the topologies used. The symbol  $\rho$  refers to the distribution models for local image regions, and  $\psi$  is the transition model between these local regions.

models of local regions [4, 86, 87, 42] to spatially variant (heterogeneous) local models [21, 5, 17, 26], depending on the assumed topology, see Fig. 5.1.

As the representational power of the model increases, it gets more complicated. At this point, global models [6, 59, 88, 89] constitute a middle ground between homogeneous local models and heterogeneous ones in terms of both computational complexity and representational power. Global priors are good at representing global features of the image space, but not at representing local features identifying the subject. As seen from Fig. 5.2-e, the results suffer from image details.

However, in constrained image domains, it is possible to approximating the performance of heterogeneous models (in representing local details) with global priors. As discussed before in Chapter 2, heterogeneous priors perform complex inference to find a good configuration of local models. In restricted domains, such as face, this flexibility is too much, because the space of possible configurations is limited. Furthermore, this constrained configuration space can be approximated with a single configuration when the sample space is arranged sufficiently compact. In other words, global models could represent image details, as long as enough local regions are used and aligned correctly. Based on this observation, we propose a global image



prior of which representational power is boosted without sacrificing its computational advantages, as shown with dashes in Fig. 5.1.

The organization of this chapter is as follows: In Section 5.1, an introduction for global models is provided through current literature. This brief review not only helps us give the basics, which constitute the main building block of our approach, but also allows discussion of the problematic assumptions that are widely used in literature. Later in Section 5.2, the details of the proposed approach are provided. Meanwhile, the remedies for the current stability problems and unrealistic assumptions are also described. In Section 5.3, the results of the experiments are evaluated in terms of representational accuracy and computational advantages. Finally concluding remarks are given in Section 5.4.

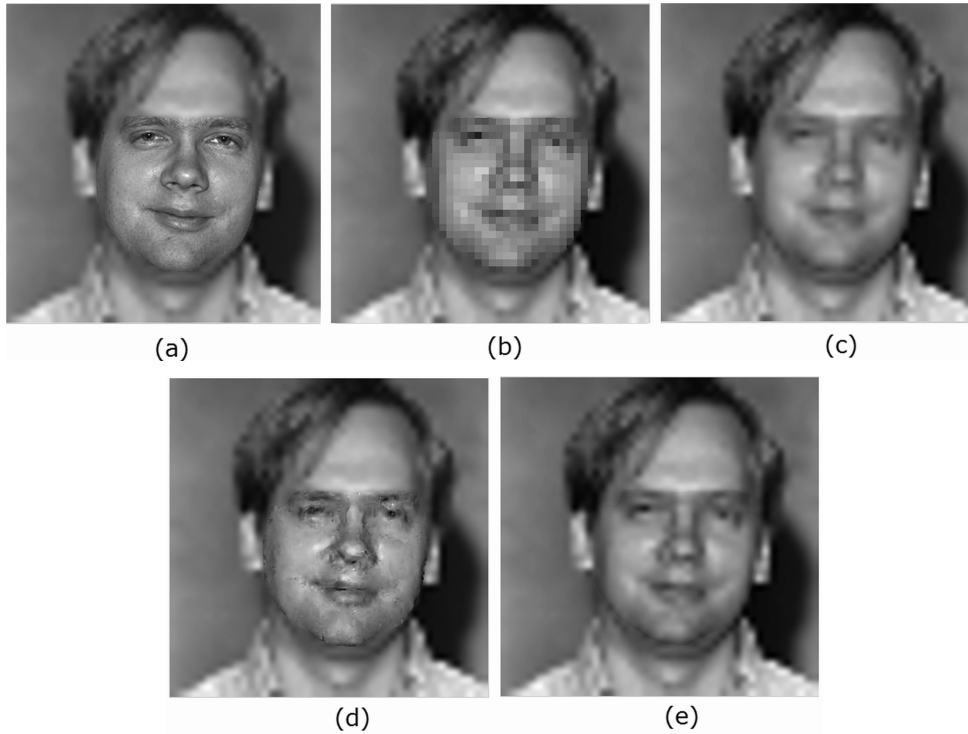
## 5.1 Background

Face Hallucination techniques attempt to reconstruct the original HR image  $I_H$  from the LR observation  $I_L$  under the assumed formation model  $I_L = HI_H + n$ , as also given in (1.7). The ML estimate of the super-resolved image is obtained by maximizing the likelihood  $\hat{I}_H \cong \arg \max_{I_H} p(I_L|I_H)$ , where it is modeled as the data constraint  $p(I_L|I_H) = p(I_L - HI_H)$  [13]. The term  $I_L - HI_H$  refers to the observation noise,  $n$  in (1.7), and the ML solution is approximated with its statistical behavior:

$$p(I_L|I_H) = \frac{1}{\sqrt{2\pi}|\Sigma_n|} \exp\left(-\frac{1}{2}(I_L - HI_H - \mu_n)\Sigma_n^{-1}(I_L - HI_H - \mu_n)^T\right). \quad (5.1)$$

Here the noise is assumed to be Gaussian  $n \sim N(\mu_n, \Sigma_n)$ , though mostly it is modeled as white Gaussian  $n \sim N(0, \sigma_n^2)$ . This highly ill-conditioned solution is regularized in the form of MAP estimation by employing a Bayesian prior model as  $\hat{I}_H \cong \arg \max_{I_H} p(I_L|I_H)p(I_H)$ .

In literature, many alternative prior designs have been proposed to describe the probability distribution function of images. For instance, homogeneous local models, such as [4, 86, 87, 42], define a first order stationary Markov Random Field (MRF) including a neighborhood prior  $p(\rho(\mathbf{\Gamma}I_{c_i}))$ , which models the spatial correlation of pixels in an image region  $c_i$ . Here  $\mathbf{\Gamma}$  is a spatial activity function, typically chosen as a derivative filter to encourage a smooth solution, and  $\rho$  is the penalty designed to be convex (such as Gaussian [4] or Huber [42]). Inference and parameter



**Figure 5.2:** Reconstruction results with different image prior models. a) HR ground-truth, b) LR observation, c) Result with spatially invariant (homogeneous) local model [4], d) Result with spatially variant (heterogeneous) local model [5], e) Result with global model [6]. Note that only the region of interest (ROI) is reconstructed and the rest is re-sampled from the noise-free LR observation.

learning are relatively simple in this modeling scheme. However the representational power is limited, because homogeneity assumption provides excessive generalization. Therefore the results suffer from over-smoothing and consist of mostly low/mid frequencies as in Fig. 5.2-c.

On the other hand, heterogeneous topologies [21, 5, 17, 26] provide more adaptive models bearing heavy computations. For each local image region, a different model is selected from a pool of candidates, learned from the training set, and the joint behavior is defined by the best model configuration. Inference of the joint model requires computationally heavy techniques, such as Belief Propagation [90], where the complexity polynomially increases as the model pool grows. Because of difficulties in modeling and inference, in real world applications only a limited number of local models can be used. Though some high-frequency content can be captured locally, this limitation causes severe global discontinuity artifacts as seen in Fig. 5.2-d.

In terms of both computational complexity and representational power, global image prior [6, 59, 88, 89] can be seen as a balanced alternative. Global priors represent the image space with a single distribution as exemplified in Fig. 5.1. This modeling scheme can be interpreted as a fixed special configuration of spatially varying local models. Based on this interpretation, it can be stated that global modeling is better than the homogeneous topologies due to employing spatially varying local models, and has less representational power than heterogeneous topologies since it allows only one configuration of local models. Rather than working on huge configuration spaces, as stated in the previous section, we think that enhancing only a single one would be more productive; especially when the computational advantages are considered.

### 5.1.1 Global image priors

Assuming that the samples constitute a Gaussian form, the global prior can be defined as:  $I_H \sim N(\mu_\rho, \Sigma_\rho)$ . In this modeling scheme, the number of parameters to be learned is reduced to the parameters of a single model, and inference can be accomplished using linear algebra.

Global modeling is also convenient to be represented in reduced dimensions [6, 59]. Working in subspaces provides several advantages, such as minimized redundancy in representation, ease of parameter learning, and increase in robustness against noise and alignment. When linear transformations are considered, image  $I$  is represented in subspace  $M$  with  $a$  based on the following relation:

$$I = Ma + \bar{I} + n_a \quad (5.2)$$

where  $n_a$  refers to the gap in representation, and  $\bar{I}$  is the mean. A popular dimensionality reduction technique, Principal Component Analysis (PCA), has been especially central to the development of face recognition algorithms [91], and to common automatic image decomposition techniques, such as the Active Appearance Model (AAM) [92]. Mathematically, a face image is represented as a linear combination of orthonormal vectors, called eigenfaces. These eigenfaces are obtained by finding the eigenvectors of the covariance matrix of the training set. When the HR image in (1.10) is transformed to a subspace by (5.2), the reconstruction turns to the

estimation of the subspace representation  $a$ :

$$\hat{a} = \arg \max_a p(I_L - HMa)p(a). \quad (5.3)$$

Similar to (1.10), the ML solution  $p(I_L - HMa)$  is modeled by the statistical behavior of the total observation gap  $v$  as:  $p(I_L - HMa) = N(v; \mu_v, \Sigma_v)$ . Here  $v$  is more than noise  $n$ , and includes also the representational gap:  $v = Hn_a + n$ .

In (5.3),  $p(a)$  refers to the transformation of the spatial image prior  $p(I_H)$ , and enforces the solution to lie on the subspace. It is simply defined as:  $a \sim N(0, \Sigma_a)$ , where the parameter  $\Sigma_a$  is designed to be the diagonal form of the component variances obtained from the PCA.

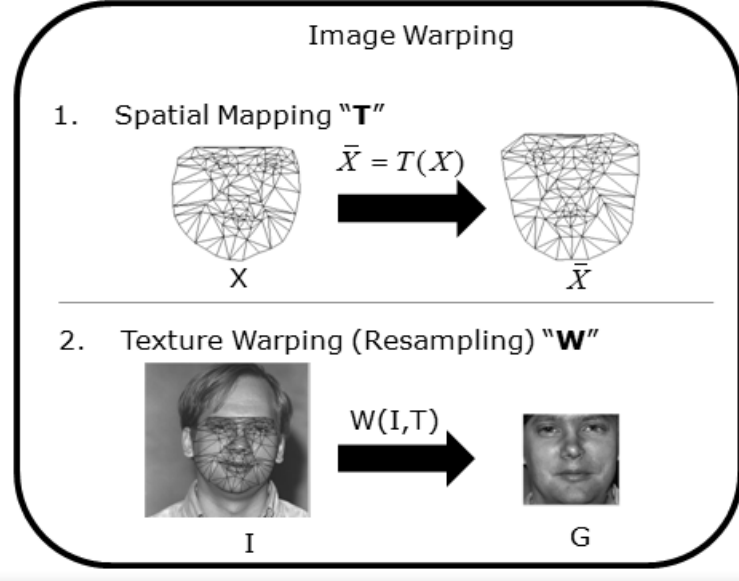
### 5.1.2 Learning model parameters

In training, the model parameters of the total residual model,  $p(v)$ , are learned from a set of HR samples,  $\{I_{H_1}, I_{H_2}, \dots, I_{H_K}\}$ . Learning starts with a preprocessing, where the training images are aligned onto a reference ground. The purpose of this preprocess is related with having a more compact sample space. The detail of alignment may change from global superimposition, such as Procrustes analysis [93], to complex shape registration [63]. As more detailed shape information is used, the accuracy in alignment increases. Therefore, the complex registration is usually selected, and the training samples are warped to a reference shape.

Image alignment consists of two consecutive processes. First, the spatial mapping is found between the shape data of the sample,  $X$ , and the reference shape,  $\bar{X}$ . Ideally, the mapping is expressed via a vector field as;  $\bar{X}(\bar{x}, \bar{y}) = X(x, y) + d(x, y)$ , where  $d(x, y)$  refers to the constant displacement at that location,  $(x, y)$ . But, it is not feasible to find the individual displacement of each pixel. Therefore, finite element discretization is applied by locally grouping the pixels as in Delaunay triangles [94]. For each triangle,  $\mathbf{t}$ , of the reference mesh,  $\bar{X}$ , a separate mapping function,  $T_t$ , is defined through its barycentric coordinates as

$$\bar{X}_t(\bar{x}, \bar{y}) = T_t(X_t(x, y)) = \sum_{i=1}^3 b_{t_i}(X_t(x, y)) \mathbf{t}_i(\bar{x}, \bar{y}), \quad (5.4)$$

where  $\mathbf{t}_i$  refers to the  $i$ th node of the triangle  $\mathbf{t}$ , and  $b_{t_i}(X_t(x, y))$  is the corresponding barycentric coordinate found on the input image [95]. Note also that  $X_t(x, y)$  and



**Figure 5.3:** Two step processing during in image warping.

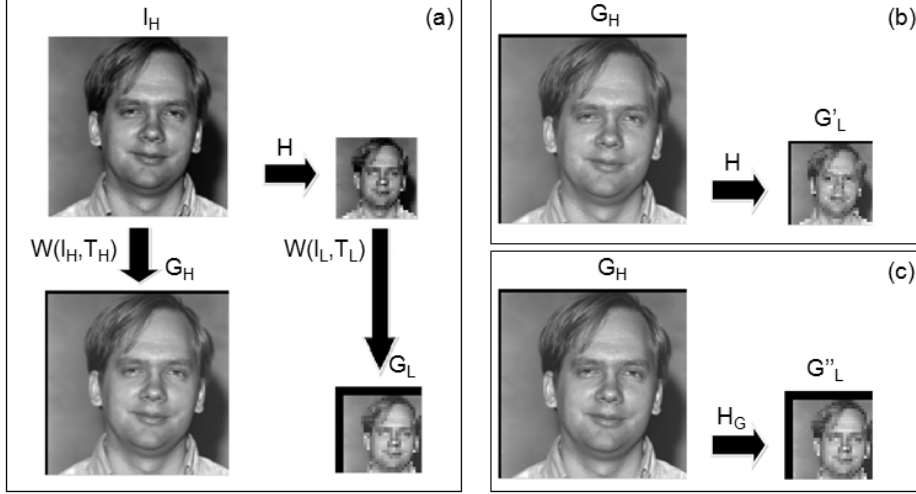
$\bar{X}_t(\bar{x}, \bar{y})$  denote the points located within the triangle  $\mathbf{t}$  of the input image and the reference ground, respectively. Considering all the triangles in  $X$ , the mapping between these two shapes can be shown as  $\bar{X} = T(X)$ .

After the spatial mapping, a re-sampling operation,  $W$ , is performed based on the spatial mapping,  $T$ , as;  $G = W(I, T)$ , where the result,  $G$ , is the alignment of  $I$ . Since all the images are aligned with the same reference shape,  $\bar{X}$ ,  $G$  is also called as the texture component of the image,  $I$ . Mostly, the bilinear interpolation is employed for the re-sampling operation,  $W$ . Assuming that a pixel in the texture,  $G(\bar{x}, \bar{y})$ , maps to  $I(x, y)$  in the original image by inverse mapping  $T_t^{-1}$ , then the corresponding pixel in the texture component,  $G(\bar{x}, \bar{y})$ , is found by

$$G(\bar{x}, \bar{y}) = w_1 I(u, v) + w_2 I(u + 1, v) + w_3 I(u, v + 1) + w_4 I(u + 1, v + 1), \quad (5.5)$$

where  $w_i$  is the weight for each neighbor, and they are found as the distances to the pixel, at  $(x, y)$ , in a quadratic way [68].

The model parameters of the residual,  $v$ , are found from the sample statistics of the aligned face images in the training set. Assuming that  $\{G_{H_1}, G_{H_2}, \dots, G_{H_K}\}$  denotes the texture components for the HR training samples,  $\{I_{H_1}, I_{H_2}, \dots, I_{H_K}\}$ , and  $\{G_{L_1}, G_{L_2}, \dots, G_{L_K}\}$  is their deformed version, obtained by the decomposition of the LR counterparts of the training images  $\{I_{L_1}, I_{L_2}, \dots, I_{L_K}\}$ , then the model parameters



**Figure 5.4:** Illustration of the biased estimate, occurring when the degradation,  $H$ , is not adapted to the alignment. In (a) HR and LR images are aligned individually as  $G_H$  and  $G_L$  (the spatial mappings,  $T_H$  and  $T_L$ , were designed as simple global translations; +2 pixels horizontal and +3 pixels vertical).  $G_L$  is the ground-truth in comparisons with (b) and (c). In (b) the degradation,  $H$ , is used in order to obtain the LR form of  $G_H$  by  $G'_L = HG_H$ . Observe that the resulting  $G'_L$  is different from the expected  $G_L$  of (a). In (c) the same operation is repeated with the corrected version of the degradation,  $H_G$ , by;  $G''_L = H_G G_H$ . Now, the result,  $G''_L$ , is the same as  $G_L$ . Note that the error in  $G'_L$  will be greater when complex spatial mappings are in question.

will be

$$\mu_v \cong \frac{1}{K} \sum_{i=1}^K (G_{L_i} - HMM^T G_{H_i}), \quad (5.6)$$

$$\Sigma_v \cong \frac{1}{K} \sum_{i=1}^K (G_{L_i} - HMM^T G_{H_i})(G_{L_i} - HMM^T G_{H_i})^T. \quad (5.7)$$

It can be observed from (5.4) and (5.5) that both the locations and intensity values of pixels are changed during alignment. Nevertheless, almost all model-based methods (except [89]) assume that the degradation operator,  $H$ , is prone to this change. They use  $H$  with the texture components,  $G_L$  and  $G_H$ , as is. However, as shown in Fig. 5.4, this would result in biased estimates because the mapping  $H$  (defined on the spatial domain between  $I_L$  and  $I_H$ ) is no more valid for the aligned images,  $G_L$  and  $G_H$ . In order to relax this dubious assumption, a correction is suggested in Section 5.2.3.

### 5.1.3 Inference and reconstruction

The reconstruction step requires the alignment of the observation so as to make it close to the sample space. Hallucination methods are divided into two main groups as to the scheme they follow for alignment. Texture-centric approaches (such as [6, 59]) make the alignment on the LR observation. They assume that the shape is preserved in resolution differences and use the LR shape information also for the HR image. With these methods, the reconstruction in (1.10) can be reached efficiently [6] by using the aforementioned quadratic models as:

$$\left[ M^T H^T \Sigma_v^{-1} H M + \Sigma_a^{-1} \right] \hat{a} = \left[ M^T H^T \Sigma_v^{-1} (G_L - \mu_v) \right]. \quad (5.8)$$

Though the estimation can be found analytically, the dubious shape preservation assumption undermines the efficiency of these methods.

Appearance-based methods have been proposed, as in [88, 89], to relax this assumption. The strategy here is based on the joint estimation of shape and texture. Thus, not only the shape preservation assumption is relaxed, but also the solution is more constrained due to the increased prior knowledge. AAM [92] may be the most popular appearance-based approach where, in addition to shape and texture models, the dependency between these components is employed by a higher PCA. In these methods, the analytical solution is sacrificed and a fitting criterion on spatial domain HR image is minimized iteratively (see Procedure 1 in Section 5.2.1).

## 5.2 Combined Model Fitting In Subspaces

The review in the preceding section reveals that utilization of shape is not easy. To overcome this problem, traditional methods either follow computationally expensive iterative processes in HR (as in [88, 89]) or make unrealistic assumptions [6, 59] to reduce the cost by sacrificing accuracy. We suggest a computationally more efficient appearance-based approach <sup>1</sup>, where the shape reconstruction is treated as a separate problem and solved in a joint framework together with texture. Moreover the stability problems and the needs of successive applications are also taken into consideration.

---

<sup>1</sup>The class name "appearance-based" is used to refer to the methods, which employ both the shape and texture components in image reconstruction.

For the realization of this idea, the images are first decomposed and represented as a combination of shape and texture components. After that, the forward and backward relations, which were previously defined in (1.7) and (1.10) for the images in the spatial domain, are re-defined individually for each component, and then transformed onto subspaces. Lastly, the resulting quadratic reconstruction expressions are optimized in a coupled way to greatly restrict the solution of each component. In the rest of this section, these steps are described in detail.

### 5.2.1 Representation of images

The highest decomposition of images results in shape  $X$  and texture  $G$  components. For LR and HR images the decompositions can be expressed with:  $I_L = \{G_L, X_L\}$  and  $I_H = \{G_H, X_H\}$ , respectively. Given the corresponding shape information  $X$ , the texture component  $G$  is extracted by image warping.

As stated in Section 5.1, global models are convenient to be represented in subspaces. Since the new global models are built over shape and texture components, they are transformed onto subspaces. Assuming that  $M_H, M_L$  and  $N_H, N_L$  denote the principal components for texture and shape, the subspace transform expressions will be:

$$G_H = M_H t_H + \bar{G}_H + e_H, \quad G_L = M_L t_L + \bar{G}_L + e_L \quad (5.9)$$

$$X_H = N_H s_H + \bar{X}_H + \varepsilon_H, \quad X_L = N_L s_L + \bar{X}_L + \varepsilon_L \quad (5.10)$$

where  $s$ 's and  $t$ 's are the new representations, and  $e$ 's and  $\varepsilon$ 's refer to the gaps. Note that for each component we use individual projections at each resolution. In this way, the components can be interpreted at their own resolutions, and this avoids the asymmetry problem. In [96] it is shown that model fitting is an asymmetric problem and in the presence of relative scaling, the warp direction ought to be chosen such that the HR image gets blurred and warped onto the LR one. Otherwise, when the input is interpreted with a model, which is not trained for input-like images, the model fitting will perform poorly.

As in our case, practical applications require making image decomposition automatically on the LR observation. We employ a gradient descent scheme, which is similar to Procedure 1, so as to automatically decompose the LR observation.



**Procedure 1.** Iterative Model Fitting proposed by Cootes et al. [92] for AAM:

1. Project the texture sample into the texture model frame using  $G_{syn} = W(I, T(Ns))$ ,
2. Evaluate the error vector,  $\mathbf{res} = G_{syn} - G_{model}$ , where  $G_{model} = Mt$ ,
3. Evaluate the current error,  $\mathbb{E} = |\mathbf{res}^2|$ ,
4. Compute the predicted displacements,  $\delta\mathbf{c} = -\Upsilon\mathbf{res}(\mathbf{c})$ ,
5. Update the model parameters  $\mathbf{c} \rightarrow \mathbf{c} + k\delta\mathbf{c}$  where initially  $k = 1$ ,
6. Calculate the new shape  $X'$  and model frame texture  $G'_{model}$ ,
7. Sample the image at the new points to obtain  $G'_{syn} = W(I, T(X'))$ ,
8. Calculate a new error vector,  $\mathbf{res}' = G'_{syn} - G'_{model}$ ,
9. If  $|\mathbf{res}'|^2 < \mathbb{E}$  then accept the new estimate; otherwise, try at  $k = 0.5$ ,  $k = 0.25$ , etc.

In this scheme, a constant linear relationship  $\Upsilon$  is assumed between the residual image,  $\mathbf{res}$ , and the additive updates. This mapping is learned offline through regression. Note also that  $\mathbf{c}$  is the subspace representation, which is obtained by a higher PCA on  $s$  and  $t$ .

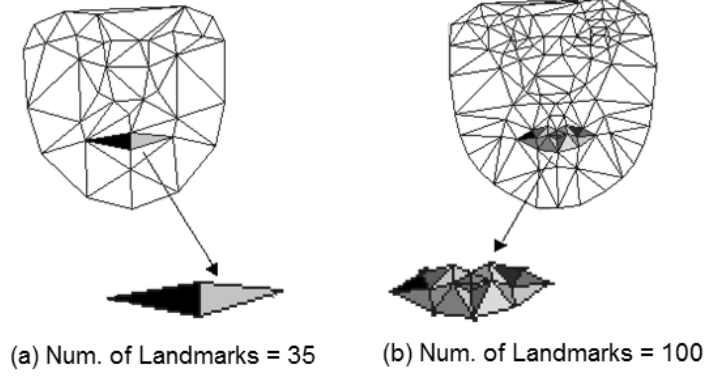
### 5.2.2 Reconstruction of shape data

Traditional model-based approaches can use only a limited number of landmarks in HR, because they obtain the shape information from the LR observation and use it for the HR image as well. However, the amount of data in LR is not enough to define a complex shape, and an accurate alignment always quests for more detailed shape information. In order to augment the shape information, the existing shape data are uniformly interpolated as in [97]. In other words, additional pseudo landmarks are arranged to be equally spaced between well defined landmark points. Though the strategy is simple, the equal spacing does not conform with the non-uniform structure of face shapes. Therefore mispositioned landmarks are produced.

In order to relax this dubious assumption and avoid inaccurate artificial landmarks, we treat the reconstruction of shape individually. The image formation and reconstruction models, given in (1.7) and (1.10), are re-defined specifically for the shape component as:

$$X_L = H_X X_H + n_X \quad (5.11)$$

$$\hat{X}_H = \arg \max_{X_H} p(X_L | X_H) p(X_H) \quad (5.12)$$



**Figure 5.5:** Illustration of the number of landmarks in the detail level of the modeling. More landmarks create more local regions which could be treated individually. In (a) the lip part can be represented with only two local models while in (b) many more models can be employed for the same region.

where the rectangular linear mapping  $H_X$  enables using shape structures in different complexities at each resolution. Note also that this individual treatment allows modeling of the deviations in shape (occurring during image deformation), and incorporating *a priori* information about the shape data in HR.

$H_X$  is simply designed as a regression operator which decimates the landmarks in HR by

$$X_L(i) = \sum_{j \in Z_H(k)} \tau_j X_H(j), \quad i \in Z_L(k) \quad \text{and} \quad \sum_{j \in Z_H(k)} \tau_j = 1, \quad (5.13)$$

where  $Z_L(k)$  and  $Z_H(k)$  refer to the landmarks related with the same image region,  $k$ , in LR and HR, respectively. For instance, let's say the lip region of a HR face is denoted by 17 landmarks, and 4 landmarks are used for the same region in LR, see Fig. 5.5. Then, the corresponding part of  $H_X$  will be a sub-matrix, which has 4 rows each consisting of 17 regression coefficients,  $\tau$ 's. The view of this part within  $H_X$  will be

$$H_X = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & 0 & \tau_{1,1} & \dots & \tau_{1,17} & 0 & \dots \\ \dots & 0 & \tau_{2,1} & \dots & \tau_{2,17} & 0 & \dots \\ \dots & 0 & \tau_{3,1} & \dots & \tau_{3,17} & 0 & \dots \\ \dots & 0 & \tau_{4,1} & \dots & \tau_{4,17} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}. \quad (5.14)$$

For each landmark in LR shape, the regression expression will be different, depending on its characteristic and the amount of available data. Especially when an excessive

amount of decimation and/or blur exists, a special  $H_X$  is required. Otherwise, when the input resolution allows for the identification of all landmarks, no rectangular operator will be required and  $H_X$  is designed traditionally as the identity matrix. Note also that  $H_X$  is valid for the whole image space, and once it is built, used by all the subjects either in training or testing.

Now these component relations, given in (5.11) and (5.12), are transformed onto a subspace. The projection of the forward model can be obtained by substituting the projections of (5.10) in (5.11) as:

$$s_L = N_L^T H_X N_H s_H + N_L^T v_S \quad (5.15)$$

where  $v_S = (H_X \varepsilon_H + n_X)$  denotes the total error in shape formation. After the backward model, in (5.12), is re-defined in terms of subspace representations as:

$$\hat{s}_H = \arg \max_{s_H} p(s_L | s_H) p(s_H). \quad (5.16)$$

Here, the ML solution  $p(s_L | s_H)$  is approximated by the probability distribution of the projected form of the total error,  $p(N_L^T v_S)$ . In (5.15),  $N_L^T v_S$  denotes the back projection error as shown below:

$$N_L^T v_S = s_L - N_L^T H_X N_H s_H. \quad (5.17)$$

Assuming that the noise has Gaussian form,  $p(v_S) \sim N(\mu_{v_S}, \Sigma_{v_S})$ , the model parameters are learned from the statistics of  $K$  samples as follows:

$$\mu_{v_S} \cong \frac{1}{K} \sum_{i=1}^K (X_{Li} - H_X N_H N_H^T X_{Hi}), \quad (5.18)$$

$$\Sigma_{v_S} \cong \frac{1}{K} \sum_{i=1}^K (X_{Li} - H_X N_H N_H^T X_{Hi})(X_{Li} - H_X N_H N_H^T X_{Hi})^T. \quad (5.19)$$

Recalling (5.15), we need the projected noise  $N_L^T v_S$  to model  $p(s_L | s_H)$ . Based on the analysis of functions of multivariate random variables [98], the projected form of the noise model is also Gaussian with the following parameters:

$$p(s_L | s_H) \sim N(N_L^T \mu_{v_S}, N_L^T \Sigma_{v_S} N_L) \quad (5.20)$$

because the eigenmatrix  $N_L$  is nonsingular, so is  $N_L^T N_L$ .

In (5.16) the regularization term,  $p(s_H)$  where  $s_H \sim N(0, \Sigma_S)$ , constrains the solution to lie on the subspace defined by the principal components (PC) of shape  $N_H$ .

With these quadratic models, the reconstruction, in (5.16), can be obtained analytically by:

$$\left[ H_S^T N_L^T \Sigma_{v_S}^{-1} N_L H_S + \Sigma_S^{-1} \right] \hat{s}_H = \left[ H_S^T N_L^T \Sigma_{v_S}^{-1} N_L (s_L - N_L^T \mu_{v_S}) \right] \quad (5.21)$$

where  $H_S = N_L^T H_X N_H$  is the projected form of the deformation.

### 5.2.3 Reconstruction of texture component

Derivations for the texture reconstruction are obtained by following a similar way to the shape reconstruction. First the image formation and reconstruction models are re-defined for texture as:

$$G_L = H_G G_H + n_G \quad (5.22)$$

$$\hat{G}_H = \arg \max_{G_H} p(G_L | G_H) p(G_H) \quad (5.23)$$

where  $G_L$  and  $G_H$  refer to the mean aligned textures corresponding to the spatial mappings  $\bar{X}_L = T_L(X_L)$  and  $\bar{X}_H = T_H(X_H)$ , respectively. In (5.22), a specific deformation operator  $H_G$  is used instead of the image deformation operator  $H$ . As explained in Section 5.1.3, warping causes change in both intensity values and locations of pixels. When  $H$  is used with the aligned textures, the least-squares solution results in biased reconstructions (see Fig. 5.4). In order to overcome this problem a correction is suggested on  $H$  by defining  $H_G$  as:

$$H_G \cong W(W(H, T_H(X_H)), T_L(X_L)) \quad (5.24)$$

where the same processing, performed in alignment, is reflected on  $H$ . Since the re-sampling in  $W$  is not linear,  $H_G$  is defined approximately. The realization of (5.24) can be made by first processing rows of  $H$  with the HR spatial mapping  $T_H$ , and then processing the columns of this intermediate result with the LR mapping  $T_L$  as summarized below.

**Procedure 2.** Correction for the deformation operator  $H$ :

1. Given the deformation operator  $H$  and the HR shape  $X_H$ ,
2. Find the corresponding  $X_L$  by using (5.11),
3. Find spatial mappings  $T_L$  and  $T_H$  for each resolution,
4. Set  $[rows\ cols] = size(H)$ , and  $H_G = 0$ ,
5. For  $i = 1$  to  $rows$ 
  - (i) Warp  $i$ 'th row of  $H$ :  $H_G(i, :) = W(H(i, :), T_H)$ ,
6. For  $j = 1$  to  $cols$ 
  - (i) Warp  $j$ 'th column of  $H_G$ :  $H_G(:, j) = W(H_G(:, j), T_L)$ ,
7. Find subspace representation of  $X_H$  via (5.10),
8. Store the sparse form of  $H_G$  together with its label  $s_H$  in a joint data structure.

The image deformation  $H$  is a highly sparse matrix consisting of the smoothing kernel parameters, which are spatially invariant. That means,  $H$  is independent from the input image and identical for the whole image space. Though  $H_G$  is also sparse, as seen in (5.24),  $H_G$  is dependent on  $X_H$ , namely  $s_H$ . A separate  $H_G$  has to be found for each intermediate solution in reconstruction. The correction in (5.24) is a costly operation,  $O(n^2)$ , since it is performed on spatial domain HR images. However, it is possible to avoid this computational burden by finding all possible  $H_G$ 's offline. We can previously calculate  $H_G$ 's for each allowed variation in  $s_H$ , and store them to use during reconstruction. Due to the limited variation in  $s_H$ , neither the number of possible  $H_G$ 's to be stored, nor the search among them would be disturbing. Furthermore, the number of  $H_G$ 's can be decreased by clustering close  $s_H$ 's and calculating only one  $H_G$  for each cluster. Fig. 5.6 reveals the need for a specific deformation operator for the reconstruction of the texture component.

After giving the models for image formation and reconstruction in the spatial domain, they are now transformed onto subspaces. The projection of the texture formation is obtained by substituting the projections of (5.9) in (5.22) as:

$$t_L = M_L^T H_G M_H t_H + M_L^T v_T \quad (5.25)$$

where  $v_T = (H_G e_H + n_G)$  denotes the total error. Similarly, the backward model, (5.23), is re-defined in subspace with:

$$\hat{t}_H = \arg \max_{t_H} p(t_L | t_H) p(t_H) \quad (5.26)$$



**Figure 5.6:** Image warping causes changes in both intensity values and locations of pixels. Therefore, the image deformation  $H$ , providing a linear mapping between the pixels at different resolutions, should be adapted to this change. Otherwise, when  $H$  is used with the aligned textures  $G_L$  and  $G_H$  in (5.22), the reconstruction in (5.23) results in biased estimates. In column (b) the LR textures  $G'_L$  have been obtained by using the image deformation operator  $H$  as:  $G'_L = HG_H$ . These textures are different from the references in column (a). In column (c), the corrected version of the deformation  $H_G$  is used to build the LR texture  $G''_L$  by:  $G''_L = H_G G_H$ . The resulting textures are almost the same with the textures in column (a).

where the ML solution,  $p(t_L|t_H)$ , is approximated with the probability distribution of the projected error  $M_L^T v_T$  as:  $p(t_L|t_H) \sim N(M_L^T \mu_{v_T}, M_L^T \Sigma_{v_T} M_L)$ . The model parameters

are learned from the statistics of  $K$  samples as follows:

$$\mu_{v_T} \cong \frac{1}{K} \sum_{i=1}^K (G_{Li} - H_G M_H M_H^T G_{Hi}) \quad (5.27)$$

$$\Sigma_{v_T} \cong \frac{1}{K} \sum_{i=1}^K (G_{Li} - H_G M_H M_H^T G_{Hi})(G_{Li} - H_G M_H M_H^T G_{Hi})^T \quad (5.28)$$

Similar to the shape prior model, the texture regularization,  $p(t_H)$  in (5.26), constrains the solution to lie on the subspace defined by the PCA model as:  $p(t_H) \sim N(0, \Sigma_T)$ .

This quadratic modeling of the reconstruction, given in (5.26), provides analytical solution by:

$$\left[ H_T^T M_L^T \Sigma_{v_T}^{-1} M_L H_T + \Sigma_T^{-1} \right] \hat{t}_H = \left[ H_T^T M_L^T \Sigma_{v_T}^{-1} M_L (t_L - M_L^T \mu_{v_T}) \right] \quad (5.29)$$

where  $H_T = M_L^T H_G M_H$ .

#### 5.2.4 Combined reconstruction

The individual reconstructions, given in (5.16) and (5.26), can be solved independently, which is common in the literature. Then it will be possible to reach the solution analytically as shown in (5.21) and (5.29). Even though this independent treatment of reconstructions provides superior reconstructions (compared to other appearance-based approaches, as shown in the experimental results in Fig. 5.10 and 5.14) [99], it is still possible to further increase the accuracy without sacrificing the linearity [72]. Since shape and texture components are statistically correlated [92], this dependency relation can be employed to better regularize the solution. The joint behavior of the image components,  $p(t_H, s_H)$ , is incorporated into the reconstruction as:

$$(\hat{t}_H, \hat{s}_H) = \arg \max_{t_H, s_H} p(t_L | t_H) p(s_L | s_H) p(t_H) p(s_H) p(t_H, s_H). \quad (5.30)$$

The models for the ML solutions ( $p(t_L | t_H)$  and  $p(s_L | s_H)$ ) and the individual priors ( $p(t_H)$  and  $p(s_H)$ ) were given previously in Sections 5.2.2 and 5.2.3. The remaining prior information  $p(t_H, s_H)$  is obtained by a higher PCA on the image components. Following the same notation with AAM [92], the correlation model is defined as:

$$\begin{bmatrix} \omega^{s_H} \\ t_H \end{bmatrix} = \begin{bmatrix} \mathbf{P} \\ \mathbf{R} \end{bmatrix} \mathbf{c} + \bar{Q} + \eta \quad (5.31)$$

where the joint principal components  $Q$  are shown in the decomposed form as:  $Q = [\mathbf{P} \ \mathbf{R}]^T$ . Also, in (5.31)  $\mathbf{c}$  is the subspace representation,  $\omega$  is the scaling,  $\bar{Q}$  is the mean (which is zero by definition since  $s_H$ 's and  $t_H$ 's are already mean normalized), and  $\eta$  is the representational gap. To express the joint behavior in terms of subspace components ( $t_H$  and  $s_H$ ), (5.31) is rearranged as:  $\mathbf{c} = \mathbf{P}^T \omega s_H + \mathbf{R}^T t_H + \eta$ . Based on this modeling, the additional prior  $p(t_H, s_H)$  is designed similarly as the constraint, enforcing the solution to lie on this joint subspace  $Q$  as:  $p(t_H, s_H) = p(\mathbf{c})$  where  $\mathbf{c} \sim N(0, \Sigma_c)$ .

Estimation in (5.30) is performed by the coupled solution of component reconstructions, given below as:

$$\hat{t}_H = \arg \max_{t_H} p(t_L | t_H) p(t_H) p(t_H, s_H) \quad (5.32)$$

$$\hat{s}_H = \arg \max_{s_H} p(s_L | s_H) p(s_H) p(t_H, s_H) \quad (5.33)$$

in a gradient descent scheme. With this sense, first the corresponding cost functions are defined by using the models given above, as:

$$\begin{aligned} \mathbb{E}(t) = & \lambda_1 (t_L - H_T t_H - M_L^T \mu_{v_T})^T (M_L^T \Sigma_{v_T} M_L)^{-1} (t_L - H_T t_H - M_L^T \mu_{v_T}) \\ & + \lambda_2 t_H^T \Sigma_t^{-1} t_H + \lambda_3 (\mathbf{P}^T \omega s_H + \mathbf{R}^T t_H)^T \Sigma_c^{-1} (\mathbf{P}^T \omega s_H + \mathbf{R}^T t_H) \end{aligned} \quad (5.34)$$

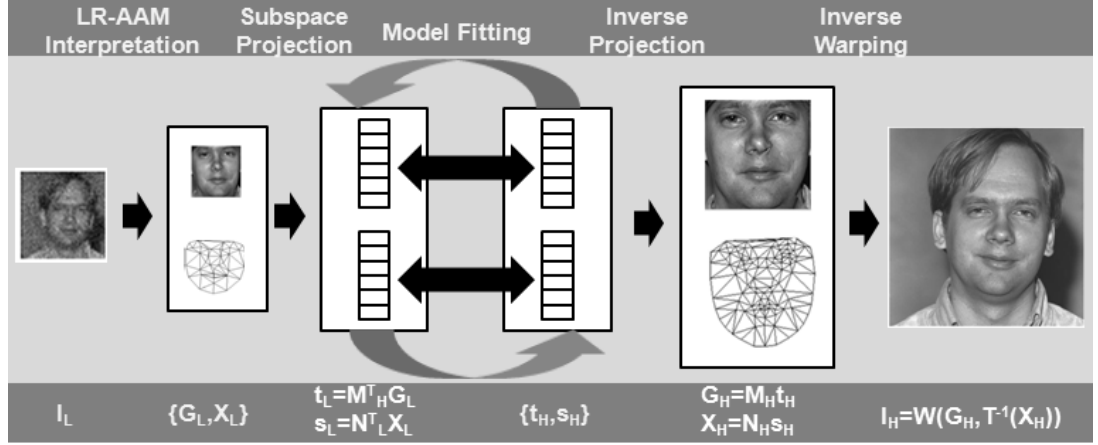
$$\begin{aligned} \mathbb{E}(s) = & \gamma_1 (s_L - H_S s_H - N_L^T \mu_{v_S})^T (N_L^T \Sigma_{v_S} N_L)^{-1} (s_L - H_S s_H - N_L^T \mu_{v_S}) \\ & + \gamma_2 s_H^T \Sigma_s^{-1} s_H + \gamma_3 (\mathbf{P}^T \omega s_H + \mathbf{R}^T t_H)^T \Sigma_c^{-1} (\mathbf{P}^T \omega s_H + \mathbf{R}^T t_H) \end{aligned}$$

where  $\lambda$  and  $\gamma$  refer to the weights adjusting the contribution of each term. The gradients of these functions will be:

$$\begin{aligned} \nabla \mathbb{E}(t) = & -\lambda_1 (H_T^T M_L^T \Sigma_{v_T}^{-1} M_L) (t_L - H_T t_H - M_L^T \mu_{v_T}) \\ & + \lambda_2 \Sigma_t^{-1} t_H + \lambda_3 (\mathbf{R} \Sigma_c^{-1}) (\mathbf{P}^T \omega s_H + \mathbf{R}^T t_H) \end{aligned} \quad (5.35)$$

$$\begin{aligned} \nabla \mathbb{E}(s) = & -\gamma_1 (H_S^T N_L^T \Sigma_{v_S}^{-1} N_L) (s_L - H_S s_H - N_L^T \mu_{v_S}) \\ & + \gamma_2 \Sigma_s^{-1} s_H + \gamma_3 (\omega \mathbf{P} \Sigma_c^{-1}) (\mathbf{P}^T \omega s_H + \mathbf{R}^T s_H). \end{aligned}$$





**Figure 5.7:** Summary of the processing followed by the proposed SR approach.

During the optimization, the image components are updated by:

$$t_H^{(n)} = t_H^{(n-1)} - \alpha^{(n)} \nabla \mathbb{E}(t_H^{(n-1)}) \quad (5.36)$$

$$s_H^{(n)} = s_H^{(n-1)} - \beta^{(n)} \nabla \mathbb{E}(s_H^{(n-1)}) \quad (5.37)$$

in the  $n$ th iteration with the step sizes  $\alpha^{(n)}$  and  $\beta^{(n)}$ . Including this coupled gradient descent optimization, the complete reconstruction process can be summarized graphically by Fig. 5.7 and procedurally as follows.

**Procedure 3.** Proposed SR reconstruction:

1. Decompose the LR observation by interpreting with the LR AAM and obtain  $s_L$  and  $t_L$ ,
2. Set  $t_H = 0$  and  $s_H = 0$ ,
3. For  $iter = 1$  to  $MaxIter$ 
  - (i) Fix  $t_H$  and estimate the new  $s_H$  via (5.37),
  - (ii) Find  $H_G$  from the repository by searching with its label  $s_H$ ,
  - (iii) Fix  $s_H$  and estimate the new  $t_H$  via (5.36),
4. Synthesize  $I_H$  from  $s_H$  and  $t_H$  by using (5.10) and (5.9), respectively.

### 5.3 Experimental Results

As shown in Fig. 5.1, defining a global image prior, which is not only powerful in representing local details but also computationally efficient, has been our intention. We performed a set of experiments to show how much the proposed solution meets this aim. The increase in the representational power has been evaluated both qualitatively and quantitatively. By using the estimations of the image components, image syntheses

have been obtained for qualitative evaluation. For quantitative comparison, the RMSE rates in subspace representations have been used. The computational advantages have been investigated by making an analysis on the running-time costs.

Results have been compared with other popular appearance-based approaches; the HR-AAM [88] and the Resolution Aware Fitting (RAF) method [89]. We refer to HR-AAM in order to describe the method, where the upscaled LR observation is interpreted with the AAM learned for HR images. In HR-AAM, the upscaled observation,  $I_U$ , is interpreted by optimizing a criterion quantifying a good match (see Procedure 1 in Section 5.3.1). In [88], the fitting criterion is given as the sum of the squared intensity differences between the intermediate synthesis and the warped observation  $W(I_U, T_L)$ , as

$$\left[ W(I_U, T(N_H s_H)) - M_{Ht_H} \right]^2. \quad (5.38)$$

Observe that the error is defined over the texture component,  $G_H$ , of the HR image,  $I_H$ , to be estimated. The results of HR-AAM get dramatically worse as the degradation increases due to the asymmetry. To overcome this problem, in the RAF algorithm [89], Dedeoglu et al. suggests a revision on the fitting criterion. In this correction, the image formation model is incorporated into the fitting criterion by

$$\left[ I_L - H(W(M_{Ht_H}, T^{-1}(N_H s_H))) \right]^2 \quad (5.39)$$

where the outcome is compared against the LR observation. Especially in severe deformations, the results of RAF would be superior to the HR-AAM results. However, the RAF algorithm is especially criticized for not utilizing statistical dependence of image components and for heavy computations [100]. At each iteration of Procedure 1, the intermediate HR synthesis is first warped back to the observation shape, and then degraded by means of  $H$ .

In our approach, computationally heavy model fitting is performed in subspaces. To allow comparison with (5.38) and (5.39), the least-squares part of the proposed solution can be given as;  $[t_L - M_L^T H_G M_{Ht_H}]^2$ , where the prior terms and the shape reconstruction have been neglected.

A set of images from the FERET database [101] has been used. The data set consists of a total of 110 different subject faces in the resolution of [360x360]. The shape

information of the images has been built by manually annotating the images with 110 landmarks. In order to create the lower resolution counterpart of the dataset, with a size of  $[45 \times 45]$ , an 8 factor decimation and blurring with a kernel (having the parameters  $N(0,3)$  and the size of  $[5 \times 5]$ ) has been applied by adding random noise with 0.005 variance for texture and 0.0001 variance for shape (noises were applied on 0-1 normalized values). The data set has been divided into two: 85 for training and 25 for testing. In order to have "shape", "texture", and "joint" principal components at each resolution, two AAMs have been trained individually for LR and HR images. Models represent 95% of the dataset and have been adjusted to search around  $+/- 3\sigma$ .

Qualitative results allow making evaluations in terms of human perception. In Fig. 5.8 we present the shape-free texture reconstructions, synthesized from the subspace estimations by using (5.9). Compared to other appearance-based methods, more realistic image details could be gained with the proposed method. The improved texture model represents the image space better, and as a result the accuracy in HR synthesis is boosted. Increased accuracy in alignment plays an important role in this improvement. To show the performance of the proposed method also in shape estimation, we present the image reconstruction results in Fig. 5.9, where the shape-free texture results in Fig. 5.8 have been warped back to the shapes estimated.

As seen from the resulting images in Fig. 5.9, the proposed method outperforms the others also in shape estimation. Reconstructions with the proposed method are realistic and close to the ground-truth. Especially the identity information of the subjects has been reconstructed more accurately. On the other hand, some reconstructions with other techniques are machinery and not like face images, though their shape-free counterparts in Fig. 5.8 are not so disturbing. This is mainly related with the inconsistency between the found shape and texture components. Recall that in the RAF algorithm, the correlation of the image components is not considered, therefore inconsistent image components could occur. This observation shows the importance of the utilization of the dependency between shape and texture.

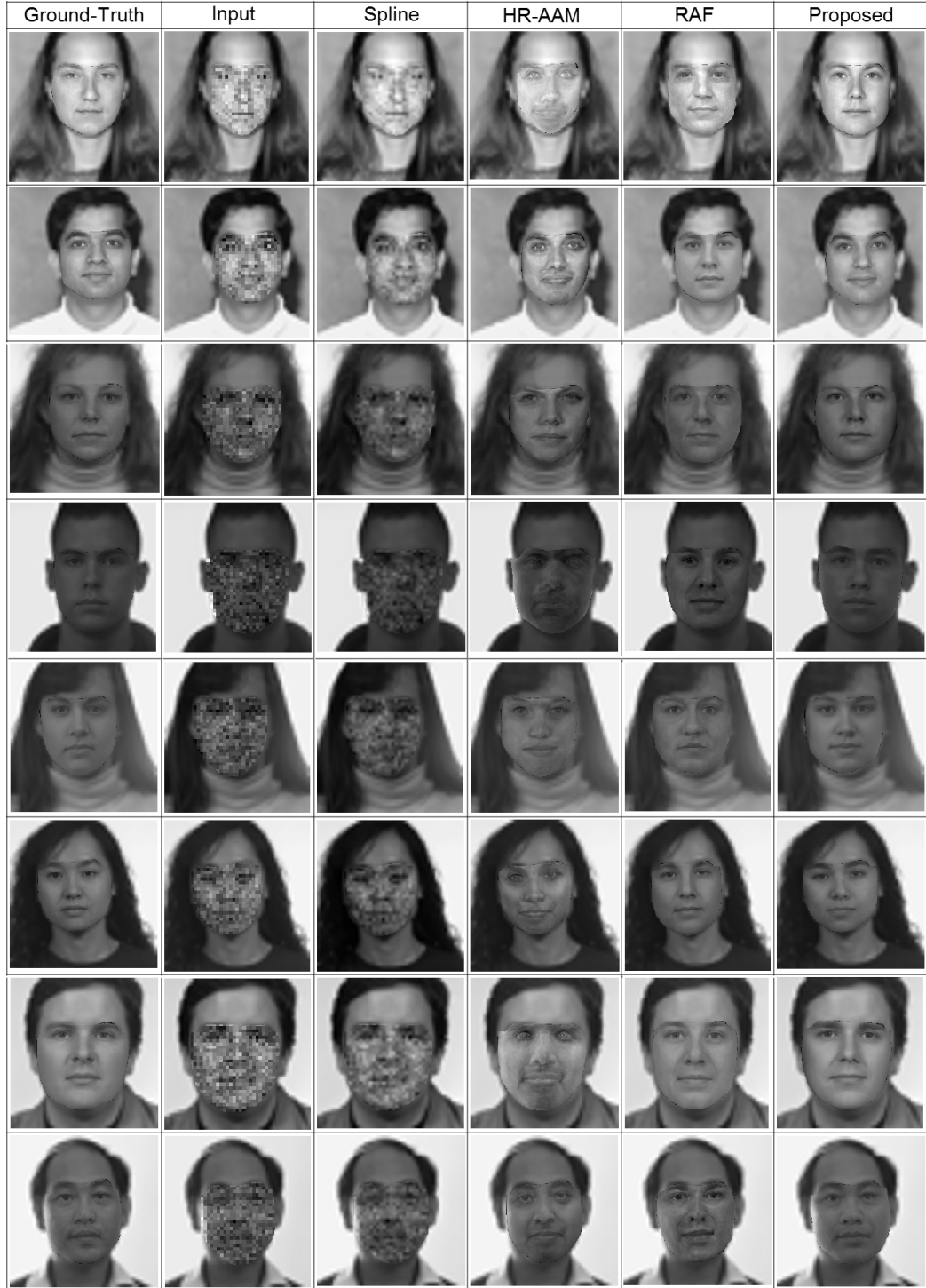
For the successive applications, such as face recognition, the accuracy in subspace representations of the image components is more important than having a visually more appealing reconstruction in spatial domain. In Fig. 5.10 and 5.11, the estimation



**Figure 5.8:** Shape-free texture syntheses.

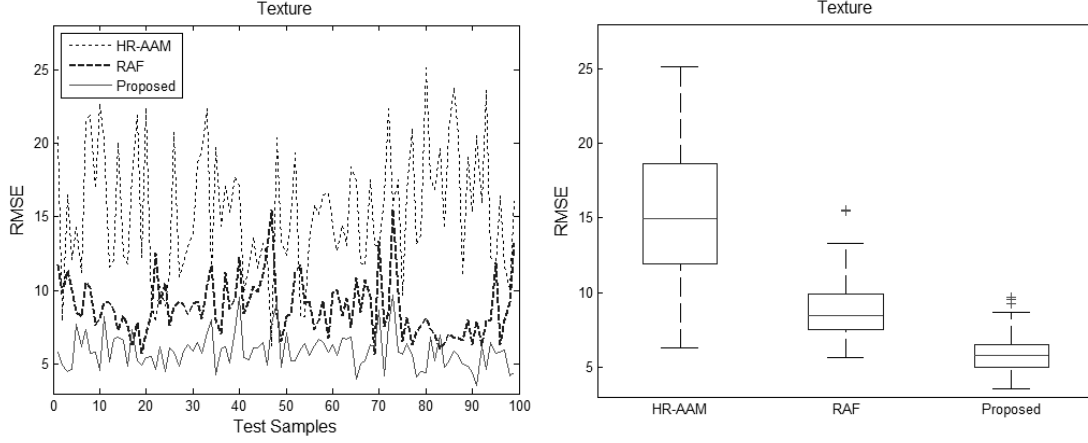
errors in subspace representations of texture and shape components are compared, respectively.

These quantitative comparisons, in Fig. 5.10 and 5.11, show that the proposed reconstruction scheme not only produces better images in spatial domain, but also significantly improves the accuracy in subspace representations. It has also been

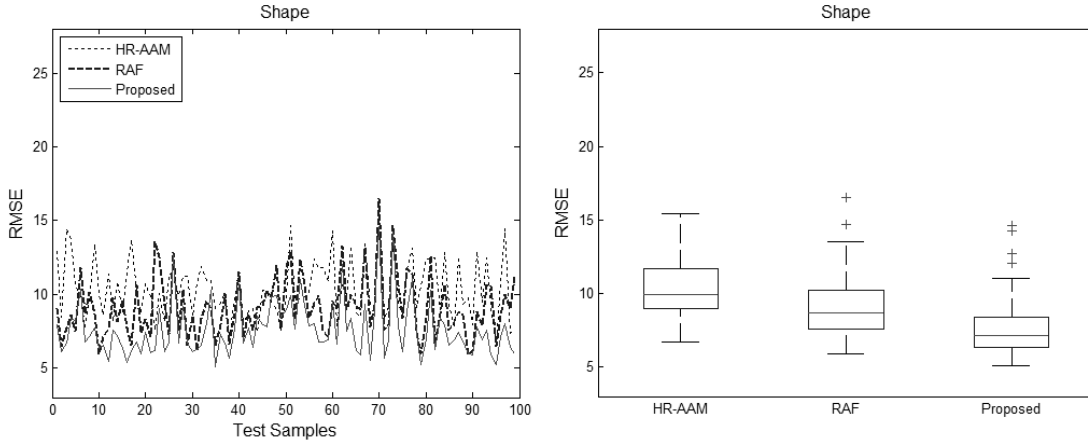


**Figure 5.9:** Qualitative results for the reconstructions from the input LR images shown in the left-most column. Note that for all columns the background has been obtained from the linear interpolation of the noise-free LR image to make the ROI more detectable.

observed that the RAF algorithm performs better than the HR-AAM due to the elimination of the asymmetry problem.



**Figure 5.10:** Root Mean Square (RMS) errors in texture subspace representations (it is also statistical summarized on the left through the box-plot representation). Note that 100 test samples have been obtained by following a leave-one-out strategy.



**Figure 5.11:** RMS errors in shape subspace representations (it is also statistical summarized on the left through the box-plot representation). Note that 100 test samples have been obtained by following a leave-one-out strategy.

Both the HR-AAM [88] and RAF [89] methods employ a fitting strategy similar to the one given in Procedure 1. This procedure has two bottlenecks with dense data. First, the re-sampling operation in Step 7 has a complexity in the order of  $O(n^2)$ , where  $n$  refers to the number of pixels in the HR image (e.g:  $n = 160 \times 160 = 25600$  in our experiments). Second, the matrix-vector multiplication, in Step 4, again has quadratic complexity,  $O(n^2)$ . Moreover the RAF algorithm includes also an additional step for image deformation.

On the other hand, the main computational load of the proposed method is caused by: interpretation of the LR input with the AAM (which is trained for LR images), and search for the right deformation  $H_G$ . They correspond to Step 1 and Step 3 in Procedure

3, respectively. As stated before, the cost of AAM interpretation with Procedure 1 is  $O(n^2)$ . Different from HR-AAM and RAF, here the model fitting is performed on the LR observation, having much fewer numbers of pixels (e.g:  $n = 20 \times 20 = 400$  in our experiments). As to Step 3, searching among  $K$  codewords can be achieved efficiently by the K-D tree algorithm [102], having a complexity in  $O(\log K)$ . Note also that the size of the codebook,  $K$ , is limited with the appearance space,  $\mathcal{Q}$ . Based on this analysis, on run-time costs, it can be claimed that the quadratic complexity of the appearance-based methods is decreased to the logarithmic complexity. Especially in severe decimations this difference is quite significant.

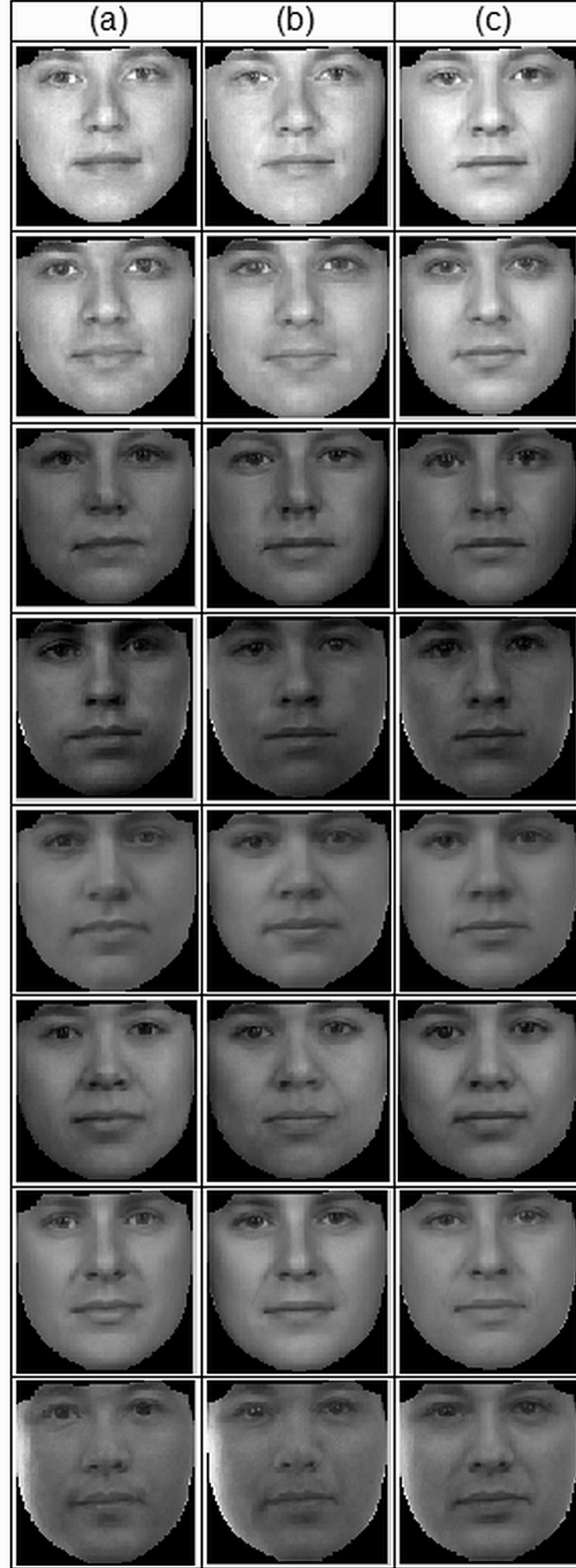
### 5.3.1 Further investigation on algorithmic details

In Fig. 5.12, we have investigated the effect of the correction, performed on the image deformation operator. The same texture reconstructions have been repeated by using the original deformation operator  $H$ , and the results have been compared with the reconstructions, obtained using the corrected version  $H_G$ .

The comparison in Fig. 5.12 shows that the textures with  $H_G$  are closer to the real texture, and include more accurate image details. Recall that a similar comparison was presented in Fig. 5.6 in Section 5.2.3. The results, in Fig. 5.6, were shown on the deterministic least-squares estimates. Therefore, the difference is more obvious in Fig. 5.6. Whereas, here in Fig. 5.12, the reconstruction results have been obtained from the MAP estimates, and some portion of the error has been absorbed by the models used for both the ML solution and prior information. In addition to these qualitative results, in Fig. 5.13 the same comparison is presented on quantitative results.

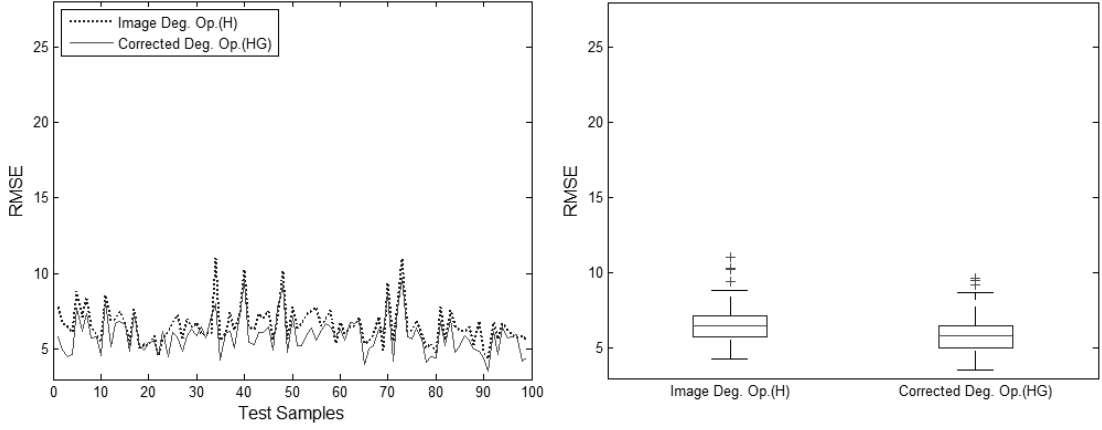
Another investigation has been performed to explore the contribution of the shape estimation in this combined strategy. The individual texture reconstruction, given in (5.29), has been performed for the same test set without caring about the shape reconstruction. In Fig. 5.14, the error rates in this experiment are compared with the results obtained by employing the combined reconstruction. As seen from this comparison, the additional constraints, coming with shape reconstruction, restrict the solution more and better reconstructions are obtained.

Up to here we have performed experiments on the synthetically deformed LR images. Different from these experiments, in Fig. 5.15, the proposed method has been

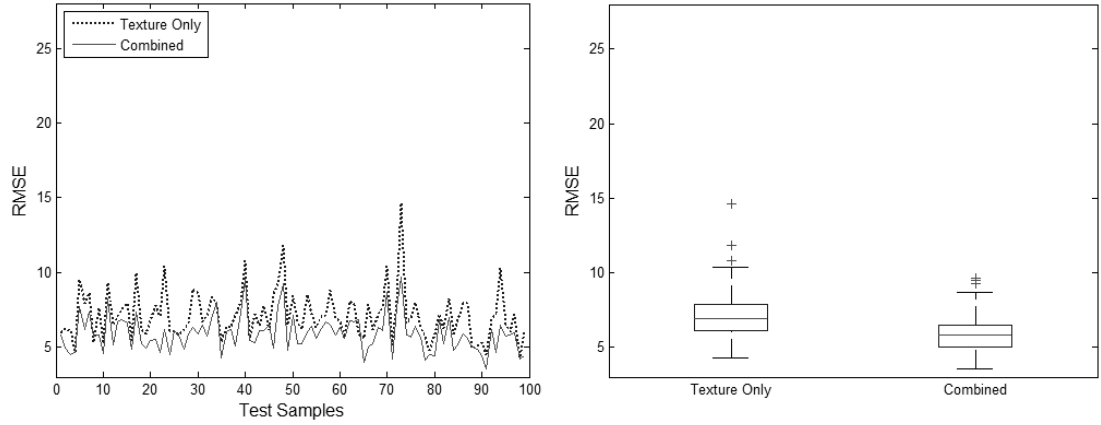


**Figure 5.12:** Effect of using the corrected form of the deformation operator on texture synthesis performance. (a) The real HR texture (b) Synthesized HR texture by using  $H_G$  in texture reconstruction (c) HR texture synthesis by using the image  $H$  in texture reconstruction.





**Figure 5.13:** RMSEs for the texture subspace representations, found by using the corrected deformation  $H_G$  in Procedure 3, compared with the error rates of the texture estimations, obtained by using the image deformation  $H$  in Procedure 3.



**Figure 5.14:** Effect of incorporation of shape into the reconstruction. Dash-line shows the error rates of the subspace representations of the texture component, which is found by employing (5.29) and neglecting shape information. Continuous line show the error rates of the texture reconstructions obtained from the combined solution proposed in Procedure 3.

tested with naturally-deformed images. For that purpose we have used the VPA Super-Resolution Face Database [103]. This new database consists of frontal face images and videos of 32 people, and it is particularly designed to test SRR techniques. The LR talking face videos were shot by a commercial SONY-DVR camera from a distance in ambient light and uncontrolled environment. The HR face images were taken by SONY-DCS F707 Digital Still Camera with closer distance again in ambient light so as to acquire face images having higher (double) resolution than those faces in the video frames. Since we perform single-frame super-resolution, we have identified the closest frames in the LR videos as the LR counterpart of the HR stills. Due to

the limited number of subjects, we have followed the leave-one-out strategy for this experiment.



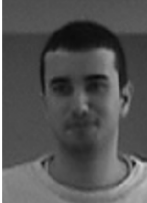
















Despite the compelling difficulties in real world scenarios (such as non-uniform noise, non-uniform and unknown PSF blur, the gap between the LR image and its HR counterpart, etc.) and the limited size of the training set, the proposed method can reconstruct much more identity information than the other techniques as seen in Fig. 5.15. Using additional image priors and modeling the residuals more accurately provide robustness to the proposed method against these difficulties.

The proposed solution framework, based on the reconstructions given in (5.32) and (5.33) as part of Procedure 3, has been fully defined in subspaces. As shown in [59, 91], subspace models are robust against noise, since they constrain the solution to lie in the face space. Moreover, to increase the robustness, obtained by subspace modeling, the noise model is learned specifically for the image space under consideration, as described in Section 5.1.1.

## **5.4 Conclusion**

Super-resolution of face images has been achieved by a new fast method, based on generative models and utilization of shape and texture components together. The major advantage of the proposed appearance-based solution is to attain the representational power of spatially varying local models while using a global model.

The representational power of the global image prior has been boosted by increasing the accuracy in the alignment and using a texture specific degradation operator. To have more accurate alignments, the shape reconstruction has been considered as a separate problem and solved in a joint framework together with texture reconstruction. This separate treatment of shape reconstruction enables both incorporating shape specific priors and using more number of landmarks with HR images. Moreover, in texture reconstruction, a specific degradation operator has been employed instead of the image degradation operator, which is originally used in image formation. Hence, the biased reconstructions, caused by using the image degradation operator with textural data, have been avoided.

Original	Input	Spline	HR-AAM	RAF	Proposed
			 $e_s=4.81$ $e_t=7.89$	 $e_s=3.63$ $e_t=5.46$	 $e_s=2.97$ $e_t=3.79$
			 $e_s=7.84$ $e_t=10.96$	 $e_s=5.31$ $e_t=5.35$	 $e_s=4.22$ $e_t=4.13$
			 $e_s=4.17$ $e_t=4.50$	 $e_s=7.97$ $e_t=8.37$	 $e_s=3.72$ $e_t=4.09$
			 $e_s=6.74$ $e_t=7.48$	 $e_s=4.11$ $e_t=7.53$	 $e_s=2.93$ $e_t=4.26$
			 $e_s=7.55$ $e_t=8.84$	 $e_s=3.16$ $e_t=6.35$	 $e_s=3.01$ $e_t=3.79$
			 $e_s=8.18$ $e_t=7.54$	 $e_s=4.83$ $e_t=6.08$	 $e_s=3.24$ $e_t=5.57$

**Figure 5.15:** Reconstruction results of naturally degraded LR observations.  $e_s$  and  $e_t$  are the RMSEs in the subspace representations of the shape and texture components, respectively. Note also that for last 3 columns the background has been obtained from the linear interpolation of the noise-free LR image to make the ROI more detectable.

The successful results in the experiments prove that the reconstructions with the proposed method have more local features constituting the identity. Moreover, the run-time cost analysis shows that the reconstruction can be obtained faster. In addition to the selected quadratic structures for modeling, the subspace transformation of the complete reconstruction expression plays a crucial role in this computational saving. In appearance-based approaches, the complexity is reduced from quadratic time to logarithmic time.

## 6. CONCLUSION & DISCUSSION

### 6.1 Summary And Contributions

This thesis presented efficient single-frame SRR methods addressing common real-world problem setups. Especially, the trade-off between the reconstruction quality and the computational complexity was focused on. It was aimed to establish the balance (stated in the Occam's principle for mathematical modeling) at the levels having high reconstruction quality and low computational cost.

Starting with the generic SRR problem definition, the exact consideration was described through the assumed forward and backward models. Meanwhile, the main characteristics of the problem were also revealed, such as: "the backward model is ill-posed", "the natural image space show heterogeneous behavior and requires adaptive treatment", "significant amount of the data is lost because of decimation" and "the applications desire the solutions to be fast, scalable and realizable with the available resources". Apparently, these characteristics define conflicting needs. For instance, as the adaptation (identifies the reconstruction quality) increases, the constraints become more complicated, and the optimization gets harder. We reviewed the related literature and identified the below list of basic principles, that should be satisfied to answer all of these needs.

- In order to maximize the information, extracted from both the observation and the reference data source, a wide set of analysis features should be employed.
- Extrapolation is difficult via learned models and limited codebooks of local image regions. Even, when smoothness is imposed, it gets much harder. So, the reference data source should be designed to have maximum textural similarity and global continuity.

- When the objective function has quadratic structure, solving mathematical programming is fastest.

In light of these principles, we developed efficient methods for different scenarios encountered frequently. The suggestions to realize these intentions are summarized below in the same ordering:

- Rather than using only the horizontal and vertical derivatives, it was suggested using a wealthier set of image features to capture more characteristics of the image space. Considering the various factors in feature set design (listed in Section 2.3) we built our proposed feature set including 1st and 2nd order derivatives at 4 intermediate orientations, and multi-scale steerable edge and bar filters. In addition to their analysis power, these features also fit well with the proposed reconstruction schemes for increased computational convenience.
- The disappointment with previous data source design attempts reveals that globally continuous realistic HF content could only be obtained by having a strong idea/experience about that content. At that point, we utilized semantically and structurally close reference/template images to represent this prior experience. Since lots of mismatches are expected, we employed robust functions while cloning the relevant details from these reference images.
- First, we proposed an iterative reconstruction scheme which does not require training and can be generalized to the natural image space. The adaptation was incorporated into the solution via the Welsch type re-descending M-estimator. Contrary to the other non-convex and discontinuous evaluation functions/estimators, the Welsch norm is convex (actually it is partially-convex; however, within the scope of this problem the initialization is generally made close to the solution. Therefore, based on this initialization assumption, it can be treated as if convex.) and has a closed form which can be differentiable up to the 2nd order. The qualitative and quantitative comparisons of the Welsch norm with the popular re-descending M-estimator, the Lorentzian norm, proved that the quality is increased significantly in addition to the computational advantages gained. Though the solution is numerical, quite satisfactory results could be reached within 30

iterations. This makes the solution promising for the real world applications needing generality.

Later, we considered the problem from the statistical learning perspective. An enhanced form of the GCRF model was utilized as the image prior. The computational conveniences of the GCRF modeling scheme made the analytical reconstruction possible. Also, through the weighting function, the adaptation of the model was increased. Thus, without sacrificing the computational advantages, we could obtain the necessary adaptive treatment for edges. Comparisons with other analytical approaches showed that a serious amount of image details could be captured.

Lastly, we addressed a more specific case where the imaging space is constrained to scenes containing only similar object/s. This restriction refers to a strong correlation between the images used in learning and the HR image to be estimated. We proposed a quite efficient method which fully utilizes this strong correlation as the image prior. Contrary to the general tendency of using local image models, global models of the shape and texture components were employed. The representational power of the global texture model was enriched with the help of shape information. For computational conveniences, convex quadratic functions were used in modeling and the variables were transformed onto subspaces. Hence, the resulting reconstruction scheme has led to quite fast algebraic operations on small-size matrices.

## **6.2 Future Directions**

There are several possible directions for future research following the present state of this study.

In our algorithms we have employed the analysis filters, which are mostly in the form of derivatives and valid for the whole natural image space. However, in the literature there are successful works, such as [104, 105, 106, 69], proposing scene-specific powerful filters. Though they are mostly used in detection applications, they can be also utilized for the SRR in addition to the generic high-pass filters. Especially for constrained domain images, they would provide better analysis performances. For

instance in [69], Torralba et al. has introduced part-based features for computer screens and cars.

Although there are works utilizing techniques from the Compressive Sensing (CS) theory, the image formation model, used in SRR, is structurally different from the formation model, which is assumed by the CS methods. In SRR the observation is not coded and directly represents the measurement data, while in CS the observation is coded by a random measurement matrix. However, the CS idea can still be employed as an artificial constraint in order to better regularize the solution. That is, this additional constraint would enforce the closeness between the coded versions of the observation and the intermediate estimate with the same measurement matrix. The randomness of the coding would probably contribute for extrapolation.

As Chapter 4 shows, the learning stage of the GCRF modeling scheme requires heavy computations, and this may cause difficulties with large-scale images. However part-based or hierarchal approaches can be adapted in order to reduce the number of unknowns. Though it would obviously require some additional processing, the efficiency of the learned models would improve. Moreover the flexible nature of the modeling scheme makes the proposed reconstruction (given in Chapter 4) quite convenient for goal-oriented purposes. The desired or expected behavior can be easily incorporated into the solution by adjusting the response estimators.



## REFERENCES

- [1] **Freeman, W., Jones, T. and Pasztor, E.**, 2000. Example-based super-resolution, *IEEE Comput. Graphi. Appl.*, **22-2**, 56–65.
- [2] **Hampel, F., Ronchetti, E. and Rousseuw, P.**, 1986. Robust Statistics - The Approach Based On Influence Functions, John Wiley & Sons, Inc., USA.
- [3] **Tappen, M., Liu, C., Adelson, E. and Freeman, W.**, 2007. Learning Gaussian Conditional Random Fields for Low-Level Vision, *Proceedings, Conf. Comp. Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1–8.
- [4] **Levitan, E. and Herman, G.**, 1987. A Maximum A Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography, *IEEE Trans. Med. Imag.*, **6-3**, 185–192.
- [5] **Liu, C., Shum, H. and Freeman, W.**, 2007. Face Hallucination: Theory and Practice, *International Journal of Computer Vision*, **75-1**, 115–134.
- [6] **Capel, D. and Zisserman, A.**, 2001. Super-Resolution from Multiple Views Using Learnt Image Models, *Proceedings, Conf. Comp. Vision and Pattern Recognition (CVPR)*, volume 2, IEEE, pp. 627–634.
- [7] **Reichenbach, S., Park, S. and Narayanswamy, R.**, 1991. Characterizing Digital Image Acquisition Devices, *Optical Engineering*, **30-2**, 170–177.
- [8] **Park, S., Park, M. and Kang, M.**, 2003. Super-resolution image reconstruction: a technical overview, *IEEE Signal Processing Magazine*, **20-3**, 21–36.
- [9] **Bishop, M.**, 2006. Pattern Recognition and Machine Learning, Springer, New York, USA.
- [10] **Baker, S. and Kanade, T.**, 2000. Hallucinating Faces, *Proceedings, Int. Conf. Automatic Face and Gesture Recognition (FG)*, IEEE, pp. 83–89.
- [11] **Tsai, R. and Huang, T.**, 1984. Advances in Computer Vision and Image Processing, *Proceeding of Inst. Elect. Eng*, **1**, 317–339.
- [12] **Lin, Z. and Shum, H.**, 2004. Fundamental limits of reconstruction based super-resolution algorithms under local translation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **26-1**, 609–616.
- [13] **Elad, M. and Feuer, A.**, 1997. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, *IEEE Trans. on Image Process.*, **6-12**, 1646–1658.

- [14] **Farsiu, S., Elad, M. and Milanfar, P.**, 2004. Fast and Robust Multi-Frame Super-resolution, *IEEE Trans. on Image Processing*, **13-10**, 1327–1344.
- [15] **Hardie, R., Barnard, K. and Armstrong, E.**, 1997. Joint MAP registration and high resolution image estimation using a sequence of under-sampled images, *IEEE Trans. Image Processing*, **6-12**, 1621–1633.
- [16] **Borman, S.**, 2004. Topics in Multiframe Superresolution Restoration, Ph.D. thesis, University of Notre Dame, Notre Dame, Indiana.
- [17] **Cheung, V., Frey, B. and Jojic, N.**, 2008. Video Epitomes, *International Journal of Computer Vision*, **76-2**, 141–152.
- [18] **Akyol, A. and Gökmen, M.**, 2008. Subspace representation of registration and reconstruction in multi-frame Super-Resolution, *Proceedings, Int. Symposium Computer and Information Sciences (ISCIS)*, IEEE, pp. 1–6.
- [19] **Woods, N., Galatsanos, N. and Katsaggelos, A.**, 2006. Stochastic Methods for Joint Registration, Restoration and Interpolation of Multiple under-sampled Images, *IEEE Trans. Image Processing*, **15-1**, 201–213.
- [20] **Fattal, R.**, 2008. Single image dehazing, *ACM Trans. Graph.*, **27-3**.
- [21] **Tappen, M., Russell, B. and Freeman, W.**, 2004. Efficient graphical models for processing images, *Proceedings, Conf. Comp. Vision and Pattern Recognition (CVPR)*, volume 2, IEEE, pp. 673–680.
- [22] **Roth, S. and Black, M.**, 2009. Fields of Experts, *International Journal of Computer Vision*, **82-2**, 205–229.
- [23] **Simoncelli, E.**, 1997. Statistical models for images: Compression, restoration and synthesis, *Proceedings, Asilomar Conference on Signals, Systems and Computers*, IEEE, pp. 673–678.
- [24] **Tappen, M., Freeman, W. and Adelson, E.**, 2005. Recovering Intrinsic Images from a Single Image, *IEEE Trans. Pattern Anal. Mach. Intell.*, **27-9**, 1459–1472.
- [25] **Jojic, N., Frey, B. and Kannan, A.**, 2003. Epitomic analysis of appearance and shape, *Proceedings, Int. Conf. Comp. Vision (ICCV)*, volume 1, IEEE, pp. 34–41.
- [26] **Yang, J., Wright, J., Huang, T. and Ma, Y.**, 2010. Image Super-Resolution Via Sparse Representation, *IEEE Transactions on Image Processing*, **19-11**, 2861–2873.
- [27] **Freeman, W., Pasztor, E. and Carmichael, O.**, 2000. Learning low-level vision, *International Journal of Computer Vision*, **40-1**, 25–47.
- [28] **Humblot, F. and Djafari, A.**, 2005. Super-resolution and joint segmentation in Bayesian framework, *Proceedings, Int. Workshop Bayesian Inference and Maximum Entropy Methods (MaxEnt)*, AIP, pp. 207–214.

- [29] **Thumblin, J. and Choudhury, P.**, 2004. Picture samples with sharp embedded boundaries, *Rendering Techniques*, 255–264.
- [30] **Battiato, S., Gallo, G. and Stanco, F.**, 2002. A locally-adaptive zooming algorithm for digital images, *Image Vision and Computing Journal*, **20-11**, 805–812.
- [31] **Patanavijit, V. and Jitapunkul, S.**, 2007. A Lorentzian Stochastic Estimation for a Robust Iterative Multiframe Super-Resolution Reconstruction with Lorentzian-Tikhonov Regularization, *EURASIP J. Adv. Sig. Proc.*
- [32] **Black, M., Sapiro, G. and Marimont, D.**, 1998. Robust anisotropic diffusion, *IEEE Transactions on Image Processing*, **7-3**, 421–432.
- [33] **Nakagaki, R. and Katsaggelos, A.**, 2003. VQ-based blind image restoration algorithm, *IEEE Trans. Image Process.*, **12-9**, 1044–1053.
- [34] **Franke, R.**, 1982. Scattered data interpolation: Tests of some methods, *Mathematics of Computation*, **38-157**, 181–200.
- [35] **Kim, S. and Bose, N.**, 1990. Reconstruction of 2-D bandlimited discrete signals from non-uniform samples, *Proceeding of Inst. Elect. Eng.*, **137**, 197–204.
- [36] **Ouwerkerk, J.**, 1984. Image super-resolution survey, *Image and Vision Computing*, **24**, 1039–1052.
- [37] **Atkins, C., Bouman, C. and Allebach, J.**, 1999. Tree-Based Resolution Synthesis, *Proceedings, Image Processing, Image Quality, Image Capture, Systems Conf. (PICS)*, volume 2, pp. 405–410.
- [38] **Atkins, C., Bouman, C. and Allebach, J.**, 2001. Optimal image scaling using pixel classification, *Proceedings, Int. Conf. Image Proc. (ICIP)*, volume 3, IEEE, pp. 864–867.
- [39] **Candocia, F. and Principe, J.**, 1999. Super-resolution of images based on local correlations, *IEEE Transactions on Neural Networks*, **10-2**, 372–380.
- [40] **Ogawa, T. and Haseyama, M.**, 2011. Adaptive Single Image Superresolution Approach Using Support Vector Data Description, *EURASIP J. Adv. Sig. Proc.*, **2011**.
- [41] **Szu, H. and Kopriva, I.**, 2001. Artificial neural networks for noisy image super-resolution, *Optics Communications*, **198-1**, 71–81.
- [42] **Schultz, R. and Stevenson, R.**, 1996. Extraction of high-resolution frames from video sequences, *IEEE Trans. Image Process.*, **5-6**, 996–1011.
- [43] **Andrews, D.F.**, 1972. Robust estimates of location: survey and advances, Princeton University Press.
- [44] **Aster, R., Borchers, B. and Thurber, C.**, 2005. Parameter Estimation and Inverse Problems, Elsevier Academic Press, Burlington, MA, USA.

- [45] **Engl, H., Hanke, M. and Neubauer, A.**, 1996. Regularization of Inverse Problems, Kluwer Academic Press, Dordrecht.
- [46] **Rubinstein, R., Zibulevsky, M. and Elad, M.**, 2010. Double sparsity: learning sparse dictionaries for sparse signal approximation, *IEEE Transactions on Signal Processing*, **58-3**, 1553–1564.
- [47] **Elad, M. and Aharon, M.**, 2006. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries, *IEEE Transactions on Image Processing*, **15-12**, 3736–3745.
- [48] **Candes, E., Romberg, J. and Tao, T.**, 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. on Information Theory*, **52-2**, 489–509.
- [49] **Chang, H., Yeung, D. and Xiong, Y.**, 2004. Super-resolution through neighbour embedding, *Proceedings, Conf. Comp. Vision and Pattern Recognition (CVPR)*, IEEE, pp. 275–282.
- [50] **Vidal, R., Ma, Y. and Sastry, S.**, 2005. Generalized Principal Component Analysis (GPCA), *IEEE Trans. Pattern Anal. Mach. Intell.*, **27-12**, 1945–1959.
- [51] **Engan, K., Aase, S. and Husoy, J.**, 1999. Frame based signal compression using method of optimal directions (MOD), *Proceedings, Int. Symposium Circuits and Systems (ISCAS)*, volume 4, IEEE, pp. 1–4.
- [52] **Zhu, S. and Mumford, D.**, 1997. Prior learning and Gibbs reaction-diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.*, **19-11**, 1236–1250.
- [53] **Haber, E. and Tenorio, L.**, 2003. Learning regularization functionals, *Inverse Problems*, **19**, 611–626.
- [54] **Hertzmann, A., Jacobs, C., Oliver, N., Curless, B. and Salesin, D.**, 2001. Image Analogies, *Proceedings, Computer Graphics and Interactive Techniques (SIGGRAPH)*, volume 28, ACM, pp. 327–340.
- [55] **Efros, A. and Leung, T.**, 1999. Texture synthesis by non-parametric sampling, *Proceedings, Comp. Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1033–1038.
- [56] **Storkey, A.**, 2003. Dynamic Structure Super-Resolution, *Proceedings, Advances in Neural Information Proc. Systems (NIPS)*, volume 15, MIT, pp. 1295–1302.
- [57] **Li, X. and Orchard, M.**, 2001. New edge-directed interpolation, *IEEE Trans. on Image Processing*, **10-10**, 1521–1527.
- [58] **Humblot, F. and Mohammad-Djafari, A.**, 2006. Super-Resolution Using Hidden Markov Model and Bayesian Detection Estimation Framework, *Journal on Applied Signal Processing*, 1–16.

- [59] **Güntürk, B., Batur, A. and Altunbaşak, Y.**, 2003. Eigenface-domain super-resolution for face recognition, *IEEE Trans. Image Process.*, **12-5**, 597–606.
- [60] **Battiato, S., Gallo, G. and Stanco, F.**, 2003. Smart interpolation by anisotropic diffusion, *Proceedings, Int. Conf. on Image Analysis and Proc. (ICIAP)*, IAPR, pp. 572–577.
- [61] **Kim, S. and Su, W.**, 1990. Recursive high-resolution reconstruction of blurred multiframe images, *IEEE Trans. Image Processing*, **2**, 534–539.
- [62] **Rhee, S. and Kang, M.**, 1999. Discrete cosine transform based regularized high-resolution image reconstruction algorithm, *Opt. Eng.*, **38-8**, 1348–1356.
- [63] **Zitova, B. and Flusser, J.**, 2003. Image registration methods: a survey, *Image Vision Comput.*, **21-11**, 977–1000.
- [64] **Akyol, A. and Gökmen, M.**, 2009. An Efficient Subspace Representation For Super-Resolution Problem, *Proceedings, Int. Conf. Image Proc., Computer Vision, & Pattern Recognition (IPCV)*, WAAS, pp. 233–242.
- [65] **El-Yamany, N. and Papamichalis, P.**, 2008. Robust Color Image Superresolution: An Adaptive M-Estimation Framework, *EURASIP J. Image and Video Processing*.
- [66] **Freeman, W. and Adelson, E.**, 1991. The Design and Use of Steerable Filters, *IEEE Trans. Pattern Anal. Mach. Intell.*, **13-9**, 891–906.
- [67] **Simoncelli, E. and Farid, H.**, 1996. Steerable Wedge Filters for Local Orientation Analysis, *IEEE Trans. Image Processing*, **5-9**, 1377–1382.
- [68] **Gonzales, R. and Woods, R.**, 2002. Digital Image Processing 2/E, Prentice Hall, New Jersey, USA.
- [69] **Torralba, A., Murphy, K. and Freeman, W.**, 2004. Sharing features: efficient boosting procedures for multiclass object detection, *Proceedings, Comp. Vision and Pattern Recognition (CVPR)*, volume 2, IEEE, pp. 762–769.
- [70] **Pennec, E. and Mallat, S.**, 2005. Sparse geometric image representations with bandelets, *IEEE Trans. Image Processing*, **14-4**, 423–438.
- [71] **Do, M. and Vetterli, M.**, 2005. The contourlet transform: an efficient directional multi-resolution image representation, *IEEE Trans. Image Processing*, **14-12**, 2091–2106.
- [72] **Akyol, A. and Gökmen, M.**, 2011. Efficient Face Hallucination By Using Shape And Texture Dependency, *Proceedings, Int. Conf. Image Proc. (ICIP)*, IEEE, pp. 1153–1156.
- [73] **Perona, P. and Malik, J.**, 1990. Scale-Space and Edge Detection Using Anisotropic Diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.*, **12-7**, 629–639.

- [74] **Tomasi, C. and Manduchi, R.**, 1998. Bilateral Filtering for Gray and Color Images, *Proceedings, Int. Conf. on Comp. Vision (ICCV)*, IEEE, pp. 839–846.
- [75] **Weijer, J.V. and den Boomgaard, R.V.**, 2001. Local mode filtering, *Proceedings, Conf. Comp. Vision and Pattern Recognition (CVPR)*, volume 2, IEEE, pp. 428–433.
- [76] **Huber, P.**, 2004. Robust Statistics, John Wiley & Sons, Inc., USA.
- [77] **Geman, S., McClure, D. and Geman, D.**, 1992. A nonlinear filter for film restoration and other problems in image processing, *CVGIP: Graphical Model and Image Processing*, **54-4**, 281–289.
- [78] **Mosteller, F. and Tukey, J.**, 1977. Data Analysis and Regression, Addison-Wesley, Reading, MA.
- [79] **Holland, P. and Welsch, R.**, 1977. Robust Regression Using Iteratively Reweighted Least-Squares, *Communications in Statistics: Theory and Methods*, **A6**, 813–827.
- [80] **Ramsay, J.**, 1977. A Comparative Study of Several Robust Estimates of Slope, Intercept, and Scale in Linear Regression, *Journal of the American Statistical Association*, **72-359**, 608–615.
- [81] **Leclerc, Y.G.**, 1989. Constructing simple stable descriptions for image partitioning, *International Journal of Computer Vision*, **3-1**, 73–102.
- [82] **Karl, W.**, 2005. Regularization in Reconstruction and Restoration, chapter Handbook of Image and Video Processing 2nd, Academic Press Limited, pp. 141–160.
- [83] **Rousseeuw, P. and Leroy, A.**, 1987. Robust Regression and Outlier Detection, Wiley, New York, USA.
- [84] **Tappen, M.**, 2007. Utilizing Variational Optimization to Learn Markov Random Fields, *Proceedings, Conf. Comp. Vision and Pattern Recognition (CVPR)*, IEEE, pp. 1–8.
- [85] **Boykov, Y., Veksler, O. and Zabih, R.**, 2001. Fast approximate energy minimization via graph cuts, *IEEE Trans. Pattern Anal. Mach. Intell.*, **23-11**, 1222–1239.
- [86] **Katsaggelos, A., Biemond, J., Schafer, R. and Mersereau, R.M.**, 1991. A regularized iterative image restoration algorithm, *IEEE Trans. Signal Process.*, **39-4**, 914–929.
- [87] **Bouman, C. and Sauer, K.**, 1993. A generalized Gaussian image model for edge-preserving MAP estimation, *IEEE Trans. Image Process.*, **2-3**, 296–310.
- [88] **Matthews, I. and Baker, S.**, 2004. Active Appearance Models Revisited, *International Journal of Computer Vision*, **60-2**, 135–164.

- [89] **Dedeoğlu, G., Baker, S. and Kanade, T.**, 2006. Resolution-Aware Fitting of Active Appearance Models to Low Resolution Images, *Proceedings, European Conf. Comp. Vision (ECCV)*, volume 2, Springer, pp. 83–97.
- [90] **Murphy, K., Weiss, Y. and Jordan, M.**, 1999. Loopy Belief Propagation for Approximate Inference: An Empirical Study, *Proceedings, Uncertainty in AI (UAI), AUAI*, pp. 467–475.
- [91] **Turk, M. and Pentland, A.**, 1991. Eigenfaces for Recognition, *Journal of Cognitive Neuro Science*, **3-2**, 71–86.
- [92] **Cootes, T., Edwards, G. and Taylor, C.**, 2001. Active Appearance Models, *IEEE Trans. Pattern Anal. Mach. Intell.*, **23-6**, 681–685.
- [93] **Kendall, D.G.**, 1989. A Survey of the Statistical Theory of Shape, *Statistical Science*, **4-2**, 87–99.
- [94] **Lertrattanapanich, S. and Bose, N.**, 2002. High resolution image formation from low resolution frames using Delaunay triangulation, *IEEE Trans. Image Process.*, **11-12**, 1427–1441.
- [95] **Bradley, C.**, 2007. The Algebra of Geometry: Cartesian, Areal and Projective Co-ordinates, Highperception, Bath, UK.
- [96] **Dedeoğlu, G., Kanade, T. and Baker, S.**, 2007. The Asymmetry of Image Registration and Its Application to Face Tracking, *IEEE Trans. Pattern Anal. Mach. Intell.*, **29-5**, 807–823.
- [97] **Cootes, T., Cooper, D., Taylor, C. and Graham, J.**, 1995. Active Shape Models - Their Training and Application, *Computer Vision and Image Understanding*, **61-1**, 38–59.
- [98] **Leon-Garcia, A.**, 2007. Probability and Random Processes For EE's (3rd Edition), Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- [99] **Akyol, A. and Gökmen, M.**, 2011. Face Super Resolution in Reduced Spaces by Using Shape and Texture, *Proceedings, Conf. Machine Vision Applications (MVA), IAPR*, pp. –.
- [100] **Akyol, A. and Gökmen, M.**, 2011. Fast Resolution Aware Model Fitting For Noisy Low Resolution Image, *Proceedings, Conf. Signal Proc. and Communications Applications (SIU), IEEE*, pp. 478–481.
- [101] **Phillips, P., Moon, H., Rizvi, S. and Rauss, P.**, 2000. The FERET Evaluation Methodology for Face Recognition Algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.*, **22-10**, 1090–1104.
- [102] **Friedman, J., Bentley, J. and Finkel, R.**, 1977. An Algorithm for Finding Best Matches in Logarithmic Expected Time, *ACM Trans. Math. Softw.*, **3-3**, 209–226.
- [103] **Sezer, O., Altunbaşak, Y. and Erçil, A.**, 2006. Face Recognition with Independent Component Based Super-resolution, *Proceedings of Visual Communications and Image Processing Conference, SPIE*.

- [104] **Zhu, S.**, 2003. Statistical Modeling and Conceptualization of Visual Patterns, *IEEE Trans. Pattern Anal. Mach. Intell.*, **25-6**, 691–712.
- [105] **Sudderth, E., Torralba, A. and Freeman, W.**, 2008. Describing Visual Scenes Using Transformed Objects and Parts, *International Journal of Computer Vision*, **77-1-3**, 291–330.
- [106] **Wu, Y., Si, Z. and Gong, H.**, 2010. Learning Active Basis Model for Object Detection and Recognition, *International Journal of Computer Vision*, **90-2**, 198–235.



## CURRICULUM VITAE

**Name Surname:** Aydın AKYOL

**Place and Date of Birth:** Afyon - April 03, 1979

**E-Mail:** akyolayd@itu.edu.tr

**B.Sc.:** Istanbul Technical University

**M.Sc.:** Sabanci University

**Professional Experience and Rewards:** Aydın Akyol received his B.Sc. degree in Control and Computer Engineering from Istanbul Technical University in 2001, and his M.Sc. degree in Computer Science from Sabanci University in 2003. In 2004 he started his Ph.D. studies in Computer Science at Istanbul Technical University. In 2007 he was a visiting scholar in the USA, where he worked with Prof. Marshall Tappen in the Computer Vision Laboratory at the University of Central Florida. During his studies, he also worked in the industry. He is the author of a journal publication and several conference papers.

### List of Publications and Patents:

- **Akyol, A.** and Erdoğan, H., 2004. Sesli diyalog sistemlerinde dolgu modeli kullanarak güvenilirlik ölçümü = Filler Model Based Confidence Measures for Dialogue Applications, *Proceedings*, Signal Proc. and Communications Applications (SIU), IEEE.
- **Akyol, A.** and Erdoğan, H. 2004., Filler model based confidence measures for spoken dialogue systems: a case study for Turkish, *Proceedings*, Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP), IEEE.
- **Akyol, A.** and Sezerman, U. 2003., Fold Classification of Protein Sequences by Genetic Algorithm, *Proceedings*, Int. Conf. Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP).
- **Akyol, A.**, Sezerman, U., Vural, E. and Işık, Z., 2002. Threading of a protein sequence to a known family fold protein, *Proceedings*, ESF Workshop on Protein Fold Prediction.
- **Akyol, A.**, 2003. Garbage Modeling in Speech Recognition, M.Sc. thesis, Sabanci University, Turkey.

## PUBLICATIONS/PRESENTATIONS ON THE THESIS

### Journal Publications:

- **Akyol, A.** and Gökmen, M., 2012. Super-Resolution reconstruction of faces by enhanced global models of shape and texture, *Pattern Recognition, Elsevier*, **45-12**, 4103-4116.

### Conference Publications:

- **Akyol, A.** and Gökmen, M., 2011. Efficient Face Hallucination by Using Shape and Texture Dependency, *Proceedings, Int. Conf. Image Proc. (ICIP)*, IEEE, pp.1153-1156.
- **Akyol, A.** and Gökmen, M., 2011. Gürültülü ve düşük çözünürlüklü imgeler için hızlı model uydurma = Fast Resolution Aware Model Fitting For Noisy Low Resolution Images, *Proceedings, Signal Proc. and Communications Applications (SIU)*, IEEE, pp.478-481.
- **Akyol, A.** and Gökmen, M., 2011. Face Super Resolution in Reduced Spaces by Using Shape and Texture, *Proceedings, Machine Vision and Applications (MVA)*, IAPR.
- **Akyol, A.** and Gökmen, M., 2009. An efficient SubSpace Representation of Super Resolution Problem, *Proceedings, Int. Conf. Image Proc. Comp. Vision (IPCV)*.
- **Akyol, A.** and Gökmen, M., 2008. SubSpace Representation of Reconstruction and Registration in SRR, *Proceedings, Int. Symposium on Comp. and Information Sciences (ISCIS)*.
- Savran, A., Çeliktutan, O., **Akyol, A.**, Trojanova, J., Dibeklioglu, H., Esenlik, S., Bozkurt, N., Demirkır, C., Akagündüz, E., Çalışkan, K., Alyüz, N., Sankur, B., Ulusoy, I., Akarun, L. and Sezgin, T.M., 2007. 3D Face Recognition Performance under Adversarial Conditions, *Proceedings, Workshop on Multi-modal Interfaces (eNTERFACE)*.
- **Akyol, A.**, Yaslan, Y. and Erol, O.K., 2007. A Genetic Programming Classifier Design Approach for Cell Images, *Proceedings, European Conf. Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*.